(12) INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)

(19) World Intellectual Property Organization
International Bureau

(43) International Publication Date
28 March 2002 (28.03.2002)

PCT

(10) International Publication Number
WO 02/24924 A2

(51) International Patent Classification^7: C12N 15/55, 9/16, 15/63, 5/10, A01K 67/027, C12P 21/00, C07K 16/40, C12Q 1/68, A61K 38/46, C12Q 1/42, G01N 33/573

(21) International Application Number: PCT/US01/29451

(22) International Filing Date: 20 September 2001 (20.09.2001)

(25) Filing Language: English

(26) Publication Language: English

(30) Priority Data:

60/234,526	21 September 2000 (21.09.2000)	US
60/236,967	29 September 2000 (29.09.2000)	US
60/238,332	5 October 2000 (05.10.2000)	US
60/242,236	19 October 2000 (19.10.2000)	US
60/243,928	27 October 2000 (27.10.2000)	US
60/249,814	17 November 2000 (17.11.2000)	US

(71) Applicant *(for all designated States except US)*: **INCYTE GENOMICS, INC.** [US/US]; 3160 Porter Drive, Palo Alto, CA 94304 (US).

(72) Inventors; and
(75) Inventors/Applicants *(for US only)*: **TANG, Y., Tom** [US/US]; 4230 Ranwick Court, San Jose, CA 95118 (US). **YAO, Monique, G.** [US/US]; 111 Frederick Court, Mountain View, CA 94043 (US). **WALIA, Narinder, K.** [US/US]; 890 Davis Street #205, San Leandro, CA 94577 (US). **ELLIOTT, Vicki, S.** [US/US]; 3770 Polton Place Way, San Jose, CA 95121 (US). **RAMKUMAR, Jayalaxmi** [IN/US]; 34359 Maybird Circle, Fremont, CA 94555 (US). **LU, Yan** [CN/US]; 3885 Corrina Way, Palo Alto, CA 94304 (US). **ARVIZU, Chandra** [US/US]; 490 Sherwood Way #1, Menlo Park, CA 94025 (US). **DING, Li** [CN/US]; 3353 Alma Street #146, Palo Alto, CA 94306 (US). **BAUGHN, Mariah, R.** [US/US]; 14244 Santiago Road, San Leandro, CA 94577 (US). **YUE, Henry** [US/US]; 826 Lois Avenue, Sunnyvale, CA 94087 (US). **LU, Dyung, Aina, M.** [US/US]; 233 Coy Drive, San Jose, CA 95123 (US). **TRIBOULEY, Catherine, M.** [FR/US]; 1121 Tennessee Street #5, San Francisco, CA 94107 (US). **THORNTON, Michael** [US/US]; 9 Medway Road, Woodside, CA 94062 (US). **GANDHI, Ameena, R.** [US/US]; 837 Roble Avenue #1, Menlo Park, CA 94025 (US). **LEE, Ernestine, A.** [US/US]; 624 Kains Street, Albany, CA 94706 (US). **XU, Yuming** [US/US]; 1739 Walnut Drive, Mountain View, CA 94040 (US). **WANG, Yumei, E.** [CN/US]; 4624 Strawberry Park Drive, San Jose, CA 95129 (US). **HAFALIA, April, J., A.** [US/US]; 2227 Calle de Primavera, Santa Clara, CA 95054 (US). **THANGAVELU, Kavitha** [IN/US]; 1950 Montecito Avenue 23, Mountain View, CA 94043 (US). **DANIELS, Susan, E.** [GB/US]; 136 Seale Avenue, Palo Alto, CA 94301 (US). **LAL, Preeti** [IN/US]; P.O. Box 5142, Santa Clara, CA 95056 (US). **SWARNAKAR, Anita** [CA/US]; 8 Locksley Avenue # 5D, San Francisco, CA 94122 (US).

(74) Agents: **HAMLET-COX, Diana** et al.; Incyte Genomics, Inc., 3160 Porter Drive, Palo Alto, CA 94304 (US).

(81) Designated States *(national)*: AE, AG, AL, AM, AT, AU, AZ, BA, BB, BG, BR, BY, BZ, CA, CH, CN, CO, CR, CU, CZ, DE, DK, DM, DZ, EC, EE, ES, FI, GB, GD, GE, GH, GM, HR, HU, ID, IL, IN, IS, JP, KE, KG, KP, KR, KZ, LC, LK, LR, LS, LT, LU, LV, MA, MD, MG, MK, MN, MW, MX, MZ, NO, NZ, PL, PT, RO, RU, SD, SE, SG, SI, SK, SL, TJ, TM, TR, TT, TZ, UA, UG, US, UZ, VN, YU, ZA, ZW.

(84) Designated States *(regional)*: ARIPO patent (GH, GM, KE, LS, MW, MZ, SD, SL, SZ, TZ, UG, ZW), Eurasian patent (AM, AZ, BY, KG, KZ, MD, RU, TJ, TM), European patent (AT, BE, CH, CY, DE, DK, ES, FI, FR, GB, GR, IE, IT, LU, MC, NL, PT, SE, TR), OAPI patent (BF, BJ, CF, CG, CI, CM, GA, GN, GQ, GW, ML, MR, NE, SN, TD, TG).

Published:
— *without international search report and to be republished upon receipt of that report*

For two-letter codes and other abbreviations, refer to the "Guidance Notes on Codes and Abbreviations" appearing at the beginning of each regular issue of the PCT Gazette.

(54) Title: PROTEIN PHOSPHATASES

(57) **Abstract:** The invention provides human protein phosphatases (PP) and polynucleotides which identify and encode PP. The invention also provides expression vectors, host cells, antibodies, agonists, and antagonists. The invention also provides for diagnosing, treating, or preventing disorders associated with aberrant expression of PP.

WO 02/24924 A2

PROTEIN PHOSPHATASES

TECHNICAL FIELD

This invention relates to nucleic acid and amino acid sequences of protein phosphatases and to the use of these sequences in the diagnosis, treatment, and prevention of immune system disorders, neurological disorders, developmental disorders, and cell proliferative disorders, including cancer, and in the assessment of the effects of exogenous compounds on the expression of nucleic acid and amino acid sequences of protein phosphatases.

BACKGROUND OF THE INVENTION

Reversible protein phosphorylation is the ubiquitous strategy used to control many of the intracellular events in eukaryotic cells. It is estimated that more than ten percent of proteins active in a typical mammalian cell are phosphorylated. Kinases catalyze the transfer of high-energy phosphate groups from adenosine triphosphate (ATP) to target proteins on the hydroxyamino acid residues serine, threonine, or tyrosine. Phosphatases, in contrast, remove these phosphate groups. Extracellular signals including hormones, neurotransmitters, and growth and differentiation factors can activate kinases, which can occur as cell surface receptors or as the activator of the final effector protein, but can also occur along the signal transduction pathway. Cascades of kinases occur, as well as kinases sensitive to second messenger molecules. This system allows for the amplification of weak signals (low abundance growth factor molecules, for example), as well as the synthesis of many weak signals into an all-or-nothing response. Phosphatases, then, are essential in determining the extent of phosphorylation in the cell and, together with kinases, regulate key cellular processes such as metabolic enzyme activity, proliferation, cell growth and differentiation, cell adhesion, and cell cycle progression.

Protein phosphatases are generally characterized as either serine/threonine- or tyrosine-specific based on their preferred phospho-amino acid substrate. However, some phosphatases (DSPs, for dual specificity phosphatases) can act on phosphorylated tyrosine, serine, or threonine residues. The protein serine/threonine phosphatases (PSPs) are important regulators of many cAMP-mediated hormone responses in cells. Protein tyrosine phosphatases (PTPs) play a significant role in cell cycle and cell signaling processes. Another family of phosphatases is the acid phosphatase or histidine acid phosphatase (HAP) family whose members hydrolyze phosphate esters at acidic pH conditions.

PSPs are found in the cytosol, nucleus, and mitochondria and in association with cytoskeletal and membranous structures in most tissues, especially the brain. Some PSPs require divalent cations, such as Ca^{2+} or Mn^{2+}, for activity. PSPs play important roles in glycogen metabolism, muscle contraction, protein synthesis, T cell function, neuronal activity, oocyte maturation, and hepatic metabolism (reviewed in Cohen, P. (1989) Annu. Rev. Biochem. 58:453-508). PSPs can be separated into two classes. The PPP class includes PP1, PP2A, PP2B/calcineurin, PP4, PP5, PP6, and PP7. Members of this class are composed of a homologous catalytic subunit bearing a very highly conserved signature sequence, coupled with one or more regulatory subunits (PROSITE PDOC00115). Further interactions with scaffold and anchoring molecules determine the intracellular localization of PSPs and substrate specificity. The PPM class consists of several closely related isoforms of PP2C and is evolutionarily unrelated to the PPP class.

PP1 dephosphorylates many of the proteins phosphorylated by cyclic AMP-dependent protein kinase (PKA) and is an important regulator of many cAMP-mediated hormone responses in cells. A number of isoforms have been identified, with the alpha and beta forms being produced by alternative splicing of the same gene. Both ubiquitous and tissue-specific targeting proteins for PP1 have been identified. In the brain, inhibition of PP1 activity by the dopamine and adenosine 3',5'-monophosphate-regulated phosphoprotein of 32kDa (DARPP-32) is necessary for normal dopamine response in neostriatal neurons (reviewed in Price, N.E. and M.C. Mumby (1999) Curr. Opin. Neurobiol. 9:336-342). PP1, along with PP2A, has been shown to limit motility in microvascular endothelial cells, suggesting a role for PSPs in the inhibition of angiogenesis (Gabel, S. et al. (1999) Otolaryngol. Head Neck Surg.121:463-468).

PP2A is the main serine/threonine phosphatase. The core PP2A enzyme consists of a single 36 kDa catalytic subunit (C) associated with a 65 kDa scaffold subunit (A), whose role is to recruit additional regulatory subunits (B). Three gene families encoding B subunits are known (PR55, PR61, and PR72), each of which contain multiple isoforms, and additional families may exist (Millward, T.A et al. (1999) Trends Biosci. 24:186-191). These "B-type" subunits are cell type- and tissue-specific and determine the substrate specificity, enzymatic activity, and subcellular localization of the holoenzyme. The PR55 family is highly conserved and bears a conserved motif (PROSITE PDOC00785). PR55 increases PP2A activity toward mitogen-activated protein kinase (MAPK) and MAPK kinase (MEK). PP2A dephosphorylates the MAPK active site, inhibiting the cell's entry into mitosis. Several proteins can compete with PR55 for PP2A core enzyme binding, including the CKII kinase catalytic subunit, polyomavirus middle and small T antigens, and SV40 small t antigen. Viruses may use this mechanism to commandeer PP2A and stimulate progression of the cell through the cell cycle (Pallas, D.C. et al. (1992) J. Virol. 66:886-893). Altered MAP kinase expression is also implicated in a variety of disease conditions including cancer, inflammation, immune disorders, and disorders affecting growth and development. PP2A, in fact, can dephosphorylate and modulate the activities of more than 30 protein kinases in vitro, and other evidence suggests that the same is true in vivo for such kinases as PKB, PKC, the calmodulin-dependent kinases, ERK family MAP kinases, cyclin-dependent kinases, and the IκB kinases (reviewed in Millward et al., supra). PP2A is itself a substrate for CKI and CKII kinases, and can be stimulated by polycationic macromolecules. A PP2A-like phosphatase is necessary to maintain the G1 phase destruction of mammalian cyclins A and B (Bastians, H. et al. (1999) Mol. Biol. Cell 10:3927-3941). PP2A is a major activity in the brain and is implicated in regulating neurofilament stability and normal neural function, particularly the phosphorylation of the microtubule-associated protein tau. Hyperphosphorylation of tau has been proposed to lead to the neuronal degeneration seen in Alzheimer's disease (reviewed in Price and Mumby, supra).

PP2B, or calcineurin, is a Ca^{2+}-activated dimeric phosphatase and is particularly abundant in the brain. It consists of catalytic and regulatory subunits, and is activated by the binding of the calcium/calmodulin complex. Calcineurin is the target of the immunosuppresant drugs cyclosporine and FK506. Along with other cellular factors, these drugs interact with calcineurin and inhibit phosphatase activity. In T cells, this blocks the calcium dependent activation of the NF-AT family of transcription factors, leading to immunosuppression. This family is widely distributed, and it is likely that calcineurin regulates gene expression in other tissues as well. In neurons, calcineurin modulates functions which range from the inhibition of neurotransmitter release to desensitization of postsynaptic NMDA-receptor coupled calcium channels to long term memory (reviewed in Price and Mumby, supra).

Other members of the PPP class have recently been identified (Cohen, P.T. (1997) Trends Biochem. Sci. 22:245-251). One of them, PP5, contains regulatory domains with tetratricopeptide repeats. It can be activated by polyunsaturated fatty acids and anionic phospholipids in vitro and appears to be involved in a number of signaling pathways, including those controlled by atrial natriuretic peptide or steroid hormones (reviewed in Andreeva, A.V. and M.A. Kutuzov (1999) Cell Signal. 11:555-562).

PP2C is a ~42kDa monomer with broad substrate specificity and is dependent on divalent cations (mainly Mn^{2+} or Mg^{2+}) for its activity. PP2C proteins share a conserved N-terminal region with an invariant DGH motif, which contains an aspartate residue involved in cation binding (PROSITE PDOC00792). Targeting proteins and mechanisms regulating PP2C activity have not been identified. PP2C has been shown to inhibit the stress-responsive p38 and Jun kinase (JNK) pathways (Takekawa, M. et al. (1998) EMBO J. 17:4744-4752).

In contrast to PSPs, tyrosine-specific phosphatases (PTPs) are generally monomeric proteins of very diverse size (from 20kDa to greater than 100kDa) and structure that function primarily in the transduction of signals across the plasma membrane. PTPs are categorized as either soluble phosphatases or transmembrane receptor proteins that contain a phosphatase domain. All PTPs share a conserved catalytic domain of about 300 amino acids which contains the active site. The active site consensus sequence includes a cysteine residue which executes a nucleophilic attack on the phosphate moiety during catalysis (Neel, B.G. and N.K. Tonks (1997) Curr. Opin. Cell Biol. 9:193-204). Receptor PTPs are made up of an N-terminal extracellular domain of variable length, a transmembrane region, and a cytoplasmic region that generally contains two copies of the catalytic domain. Although only the first copy seems to have enzymatic activity, the second copy apparently affects the substrate specificity of the first. The extracellular domains of some receptor PTPs contain fibronectin-like repeats, immunoglobulin-like domains, MAM domains (an extracellular motif likely to have an adhesive function), or carbonic anhydrase-like domains (PROSITE PDOC 00323). This wide variety of structural motifs accounts for the diversity in size and specificity of PTPs.

PTPs play important roles in biological processes such as cell adhesion, lymphocyte activation, and cell proliferation. PTPs μ and κ are involved in cell-cell contacts, perhaps regulating cadherin/catenin function. A number of PTPs affect cell spreading, focal adhesions, and cell motility, most of them via the integrin/tyrosine kinase signaling pathway (reviewed in Neel and Tonks, supra). CD45 phosphatases regulate signal transduction and lymphocyte activation (Ledbetter, J.A. et al. (1988) Proc. Natl. Acad. Sci. USA 85:8628-8632). Soluble PTPs containing Src-homology-2 domains have been identified (SHPs), suggesting that these molecules might interact with receptor tyrosine kinases. SHP-1 regulates cytokine receptor signaling by controlling the Janus family PTKs in hematopoietic cells, as well as signaling by the T-cell receptor and c-Kit (reviewed in Neel and Tonks, supra). M-phase inducer phosphatase plays a key role in the induction of mitosis by dephosphorylating and activating the PTK CDC2, leading to cell division (Sadhu, K. et al. (1990) Proc. Natl. Acad. Sci. USA 87:5139-5143). In addition, the genes encoding at least eight PTPs have been mapped to chromosomal regions that are translocated or rearranged in various neoplastic conditions, including lymphoma, small cell lung carcinoma, leukemia, adenocarcinoma, and neuroblastoma (reviewed in Charbonneau, H. and N.K. Tonks (1992) Annu. Rev. Cell Biol. 8:463-493). The PTP enzyme active site comprises the consensus sequence of the MTM1 gene family. The MTM1 gene is responsible for X-linked recessive myotubular myopathy, a congenital muscle disorder that has been linked to Xq28 (Kioschis, P. et al., (1998) Genomics 54:256-266. Myotubularin is a PTP which is required for muscle differentiation and is a potent phospatidylinositol 3-phosphate (PI(3)P) phosphatase. Mutations in the myotubularin gene (MTM1) that cause human myotubular myopathy result in a dramatic reduction in the ability of the phosphatase to dephosphorylate PI(3)P. Myotubular myopathy is an X-linked, severe congenital disorder characterized by generalized muscle weakness and impaired maturation of muscle fibers (Taylor,G.S. et al., (2000) Proc. Natl. Acad. Sci. U.S.A. 97:8910-8915). Many PTKs are encoded by oncogenes, and it is well known that oncogenesis is often accompanied by increased tyrosine phosphorylation activity. It is therefore possible that PTPs may serve to prevent or reverse cell transformation and the growth of various cancers by controlling the levels of tyrosine phosphorylation in cells. This is supported by studies showing that overexpression of PTP can suppress transformation in cells and that specific inhibition of PTP can enhance cell transformation (Charbonneau and Tonks, supra).

Dual specificity phosphatases (DSPs) are structurally more similar to the PTPs than the PSPs. DSPs bear an extended PTP active site motif with an additional 7 amino acid residues. DSPs are primarily associated with cell proliferation and include the cell cycle regulators cdc25A, B, and C. The phosphatases DUSP1 and DUSP2 inactivate the MAPK family members ERK (extracellular signal-regulated kinase), JNK (c-Jun N-terminal kinase), and p38 on both tyrosine and threonine residues (PROSITE PDOC 00323, supra). In the activated state, these kinases have been implicated in neuronal differentiation, proliferation, oncogenic transformation, platelet aggregation, and apoptosis. Thus, DSPs are necessary for proper regulation of these processes (Muda, M. et al. (1996) J. Biol. Chem. 271:27205-27208). The tumor suppressor PTEN is a DSP that also shows lipid phosphatase activity. It seems to negatively regulate interactions with the extracellular matrix and maintains sensitivity to apoptosis. PTEN has been implicated in the prevention of angiogenesis (Giri, D. and M. Ittmann (1999) Hum. Pathol. 30:419-424) and abnormalities in its expression are associated with numerous cancers (reviewed in Tamura, M. et al. (1999) J. Natl. Cancer Inst. 91:1820-1828).

Histidine acid phosphatase (HAP; EXPASY EC 3.1.3.2), also known as acid phosphatase, hydrolyzes a wide spectrum of substrates including alkyl, aryl, and acyl orthophosphate monoesters and phosphorylated proteins at low pH. HAPs share two regions of conserved sequences, each centered around a histidine residue which is involved in catalytic activity. Members of the HAP family include lysosomal acid phosphatase (LAP) and prostatic acid phosphatase (PAP), both sensitive to inhibition by L-tartrate (PROSITE PDOC00538).

LAP, an orthophosphoric monoester of the endosomal/lysosomal compartment is a housekeeping gene whose enzymatic activity has been detected in all tissues examined (Geier, C. et al. (1989) Eur. J. Biochem. 183:611-616). LAP-deficient mice have progressive skeletal disorder and an increased disposition toward generalized seizures (Saftig, P. et al. (1997) J. Biol. Chem. 272:18628-18635). LAP-deficient patients were found to have the following clinical features: intermittent vomiting, hypotonia, lethargy, opisthotonos, terminal bleeding, seizures, and death in early infancy (Online Mendelian Inheritance in Man (OMIM) *200950).

PAP, a prostate epithelium-specific differentiation antigen produced by the prostate gland, has been used to diagnose and stage prostate cancer. In prostate carcinomas, the enzymatic activity of PAP was shown to be decreased compared with normal or benign prostate hypertrophy cells (Foti, A.G. et al. (1977) Cancer Res. 37:4120-4124). Two forms of PAP have been identified, secreted and intracellular. Mature secreted PAP is detected in the seminal fluid and is active as a glycosylated homodimer with a molecular weight of approximately 100-kilodalton. Intracellular PAP is found to exhibit endogenous phosphotyrosyl protein phosphatase activity and is involved in regulating prostate cell growth (Meng, T.C. and M.F. Lin (1998) J. Biol. Chem. 34:22096-22104).

Synaptojanin, a polyphosphoinositide phosphatase, dephosphorylates phosphoinositides at positions 3, 4 and 5 of the inositol ring. Synaptojanin is a major presynaptic protein found at clathrin-coated endocytic intermediates in nerve terminals, and binds the clathrin coat-associated protein, EPS15. This binding is mediated by the C-terminal region of synatojanin-170, which has 3 Asp-Pro-Phe amino acid repeats. Further, this 3 residue repeat had been found to be the binding site for the EH domains of EPS15 (Haffner, C. et al. (1997) FEBS Lett. 419:175-180). Additionally, synaptojanin may potentially regulate interactions of endocytic proteins with the plasma membrane, and be involved in synaptic vesicle recycling (Brodin, L. et al. (2000) Curr. Opin. Neurobiol. 10:312-320). Studies in mice with a targeted disruption in the synaptojanin 1 gene (Synj1) were shown to support coat formation of endocytic vesicles more effectively than was seen in wild-type mice, suggesting that Synj1 can act as a negative regulator of membrane-coat protein interactions. These findings provide genetic evidence for a crucial role of phosphoinositide metabolism in synaptic vesicle recycling (Cremona, O. et al. (1999) Cell 99:179-188).

The discovery of new protein phosphatases, and the polynucleotides encoding them, satisfies a need in the art by providing new compositions which are useful in the diagnosis, prevention, and treatment of immune system disorders, neurological disorders, developmental disorders, and cell proliferative disorders, including cancer, and in the assessment of the effects of exogenous compounds on the expression of nucleic acid and amino acid sequences of protein phosphatases.

SUMMARY OF THE INVENTION

The invention features purified polypeptides, protein phosphatases, referred to collectively as "PP" and individually as "PP-1," "PP-2," "PP-3," "PP-4," "PP-5," "PP-6," "PP-7," "PP-8," "PP-9," "PP-10," "PP-11" and "PP-12." In one aspect, the invention provides an isolated polypeptide selected from the group consisting of a) a polypeptide comprising an amino acid sequence selected from the group consisting of SEQ ID NO:1-12, b) a polypeptide comprising a naturally occurring amino acid sequence at least 90% identical to an amino acid sequence selected from the group consisting of SEQ ID NO:1-12, c) a biologically active fragment of a polypeptide having an amino acid sequence selected from the group consisting of SEQ ID NO:1-12, and d) an immunogenic fragment of a polypeptide having an amino acid sequence selected from the group consisting of SEQ ID NO:1-12. In one alternative, the invention provides an isolated polypeptide comprising the amino acid sequence of SEQ ID NO:1-12.

The invention further provides an isolated polynucleotide encoding a polypeptide selected from the group consisting of a) a polypeptide comprising an amino acid sequence selected from the group consisting of SEQ ID NO:1-12, b) a polypeptide comprising a naturally occurring amino acid sequence at least 90% identical to an amino acid sequence selected from the group consisting of SEQ ID NO:1-12, c) a biologically active fragment of a polypeptide having an amino acid sequence selected from the group consisting of SEQ ID NO:1-12, and d) an immunogenic fragment of a polypeptide having an amino acid sequence selected from the group consisting of SEQ ID NO:1-12. In one alternative, the polynucleotide encodes a polypeptide selected from the group consisting of SEQ ID NO:1-12. In another alternative, the polynucleotide is selected from the group consisting of SEQ ID NO:13-24.

Additionally, the invention provides a recombinant polynucleotide comprising a promoter sequence operably linked to a polynucleotide encoding a polypeptide selected from the group consisting of a) a polypeptide comprising an amino acid sequence selected from the group consisting of SEQ ID NO:1-12, b) a polypeptide comprising a naturally occurring amino acid sequence at least 90% identical to an amino acid sequence selected from the group consisting of SEQ ID NO:1-12, c) a biologically active fragment of a polypeptide having an amino acid sequence selected from the group consisting of SEQ ID NO:1-12, and d) an immunogenic fragment of a polypeptide having an amino acid sequence selected from the group consisting of SEQ ID NO:1-12. In one alternative, the invention provides a cell transformed with the recombinant polynucleotide. In another alternative, the invention provides a transgenic organism comprising the recombinant polynucleotide.

The invention also provides a method for producing a polypeptide selected from the group consisting of a) a polypeptide comprising an amino acid sequence selected from the group consisting of SEQ ID NO:1-12, b) a polypeptide comprising a naturally occurring amino acid sequence at least 90% identical to an amino acid sequence selected from the group consisting of SEQ ID NO:1-12, c) a biologically active fragment of a polypeptide having an amino acid sequence selected from the group consisting of SEQ ID NO:1-12, and d) an immunogenic fragment of a polypeptide having an amino acid sequence selected from the group consisting of SEQ ID NO:1-12. The method comprises a) culturing a cell under conditions suitable for expression of the polypeptide, wherein said cell is transformed with a recombinant polynucleotide comprising a promoter sequence operably linked to a polynucleotide encoding the polypeptide, and b) recovering the polypeptide so expressed.

Additionally, the invention provides an isolated antibody which specifically binds to a polypeptide selected from the group consisting of a) a polypeptide comprising an amino acid sequence selected from the group consisting of SEQ ID NO:1-12, b) a polypeptide comprising a naturally occurring amino acid sequence at least 90% identical to an amino acid sequence selected from the group consisting of SEQ ID NO:1-12, c) a biologically active fragment of a polypeptide having an amino acid sequence selected from the group consisting of SEQ ID NO:1-12, and d) an immunogenic fragment of a polypeptide having an amino acid sequence selected from the group consisting of SEQ ID NO:1-12.

The invention further provides an isolated polynucleotide selected from the group consisting of a) a polynucleotide comprising a polynucleotide sequence selected from the group consisting of SEQ ID NO:13-24, b) a polynucleotide comprising a naturally occurring polynucleotide sequence at least 90% identical to a polynucleotide sequence selected from the group consisting of SEQ ID NO:13-24, c) a polynucleotide complementary to the polynucleotide of a), d) a polynucleotide complementary to the polynucleotide of b), and e) an RNA equivalent of a)-d). In one alternative, the polynucleotide comprises at least 60 contiguous nucleotides.

Additionally, the invention provides a method for detecting a target polynucleotide in a sample, said target polynucleotide having a sequence of a polynucleotide selected from the group consisting of a) a polynucleotide comprising a polynucleotide sequence selected from the group consisting of SEQ ID NO:13-24, b) a polynucleotide comprising a naturally occurring polynucleotide sequence at least 90% identical to a polynucleotide sequence selected from the group consisting of SEQ ID NO:13-24, c) a polynucleotide complementary to the polynucleotide of a), d) a polynucleotide complementary to the polynucleotide of b), and e) an RNA equivalent of a)-d). The method comprises a) hybridizing the sample with a probe comprising at least 20 contiguous nucleotides comprising a sequence complementary to said target polynucleotide in the sample, and which probe specifically hybridizes to said target polynucleotide, under conditions whereby a hybridization complex is formed between said probe and said target polynucleotide or fragments thereof, and b) detecting the presence or absence of said hybridization complex, and optionally, if present, the amount thereof. In one alternative, the probe comprises at least 60 contiguous nucleotides.

The invention further provides a method for detecting a target polynucleotide in a sample, said target polynucleotide having a sequence of a polynucleotide selected from the group consisting of a) a polynucleotide comprising a polynucleotide sequence selected from the group consisting of SEQ ID NO:13-24, b) a polynucleotide comprising a naturally occurring polynucleotide sequence at least 90% identical to a polynucleotide sequence selected from the group consisting of SEQ ID NO:13-24, c) a polynucleotide complementary to the polynucleotide of a), d) a polynucleotide complementary to the polynucleotide of b), and e) an RNA equivalent of a)-d). The method comprises a) amplifying said target polynucleotide or fragment thereof using polymerase chain reaction amplification, and b) detecting the presence or absence of said amplified target polynucleotide or fragment thereof, and, optionally, if present, the amount thereof.

The invention further provides a composition comprising an effective amount of a polypeptide selected from the group consisting of a) a polypeptide comprising an amino acid sequence selected from the group consisting of SEQ ID NO:1-12, b) a polypeptide comprising a naturally occurring amino acid sequence at least 90% identical to an amino acid sequence selected from the group consisting of SEQ ID NO:1-12, c) a biologically active fragment of a polypeptide having an amino acid sequence selected from the group consisting of SEQ ID NO:1-12, and d) an immunogenic fragment of a polypeptide having an amino acid sequence selected from the group consisting of SEQ ID NO:1-12, and a pharmaceutically acceptable excipient. In one embodiment, the composition comprises an amino acid sequence selected from the group consisting of SEQ ID NO:1-12. The invention additionally provides a method of treating a disease or condition associated with decreased expression of functional PP, comprising administering to a patient in need of such treatment the composition.

The invention also provides a method for screening a compound for effectiveness as an agonist of a polypeptide selected from the group consisting of a) a polypeptide comprising an amino acid sequence selected from the group consisting of SEQ ID NO:1-12, b) a polypeptide comprising a naturally occurring amino acid sequence at least 90% identical to an amino acid sequence selected from the group consisting of SEQ ID NO:1-12, c) a biologically active fragment of a polypeptide having an amino acid sequence selected from the group consisting of SEQ ID NO:1-12, and d) an immunogenic fragment of a polypeptide having an amino acid sequence selected from the group consisting of SEQ ID NO:1-12. The method comprises a) exposing a sample comprising the polypeptide to a compound, and b) detecting agonist activity in the sample. In one alternative, the invention provides a composition comprising an agonist compound identified by the method and a pharmaceutically acceptable excipient. In another alternative, the invention provides a method of treating a disease or condition associated with decreased expression of functional PP, comprising administering to a patient in need of such treatment the composition.

Additionally, the invention provides a method for screening a compound for effectiveness as an antagonist of a polypeptide selected from the group consisting of a) a polypeptide comprising an amino acid sequence selected from the group consisting of SEQ ID NO:1-12, b) a polypeptide comprising a naturally occurring amino acid sequence at least 90% identical to an amino acid sequence selected from the group consisting of SEQ ID NO:1-12, c) a biologically active fragment of a polypeptide having an amino acid sequence selected from the group consisting of SEQ ID NO:1-12, and d) an immunogenic fragment of a polypeptide having an amino acid sequence selected from the group consisting of SEQ ID NO:1-12. The method comprises a) exposing a sample comprising the polypeptide to a compound, and b) detecting antagonist activity in the sample. In one alternative, the invention provides a composition comprising an antagonist compound identified by the method and a pharmaceutically acceptable excipient. In another alternative, the invention provides a method of treating a disease or condition associated with overexpression of functional PP, comprising administering to a patient in need of such treatment the composition.

The invention further provides a method of screening for a compound that specifically binds to a polypeptide selected from the group consisting of a) a polypeptide comprising an amino acid sequence selected from the group consisting of SEQ ID NO:1-12, b) a polypeptide comprising a naturally occurring amino acid sequence at least 90% identical to an amino acid sequence selected from the group consisting of SEQ ID NO:1-12, c) a biologically active fragment of a polypeptide having an amino acid sequence selected from the group consisting of SEQ ID NO:1-12, and d) an immunogenic fragment of a polypeptide having an amino acid sequence selected from the group consisting of SEQ ID NO:1-12. The method comprises a) combining the polypeptide with at least one test compound under suitable conditions, and b) detecting binding of the polypeptide to the test compound, thereby identifying a compound that specifically binds to the polypeptide.

The invention further provides a method of screening for a compound that modulates the activity of a polypeptide selected from the group consisting of a) a polypeptide comprising an amino acid sequence selected from the group consisting of SEQ ID NO:1-12, b) a polypeptide comprising a naturally occurring amino acid sequence at least 90% identical to an amino acid sequence selected from the group consisting of SEQ ID NO:1-12, c) a biologically active fragment of a polypeptide having an amino acid sequence selected from the group consisting of SEQ ID NO:1-12, and d) an immunogenic fragment of a polypeptide having an amino acid sequence selected from the group consisting of SEQ ID NO:1-12. The method comprises a) combining the polypeptide with at least one test compound under conditions permissive for the activity of the polypeptide, b) assessing the activity of the polypeptide in the presence of the test compound, and c) comparing the activity of the polypeptide in the presence of the test compound with the activity of the polypeptide in the absence of the test compound, wherein a change in the activity of the polypeptide in the presence of the test compound is indicative of a compound that modulates the activity of the polypeptide.

The invention further provides a method for screening a compound for effectiveness in altering expression of a target polynucleotide, wherein said target polynucleotide comprises a polynucleotide sequence selected from the group consisting of SEQ ID NO:13-24, the method comprising a) exposing a sample comprising the target polynucleotide to a compound, and b) detecting altered expression of the target polynucleotide.

The invention further provides a method for assessing toxicity of a test compound, said method comprising a) treating a biological sample containing nucleic acids with the test compound; b) hybridizing the nucleic acids of the treated biological sample with a probe comprising at least 20 contiguous nucleotides of a polynucleotide selected from the group consisting of i) a polynucleotide comprising a polynucleotide sequence selected from the group consisting of SEQ ID NO:13-24, ii) a polynucleotide comprising a naturally occurring polynucleotide sequence at least 90% identical to a polynucleotide sequence selected from the group consisting of SEQ ID NO:13-24, iii) a polynucleotide having a sequence complementary to i), iv) a polynucleotide complementary to the polynucleotide of ii), and v) an RNA equivalent of i)-iv). Hybridization occurs under conditions whereby a specific hybridization complex is formed between said probe and a target polynucleotide in the biological sample, said target polynucleotide selected from the group consisting of i) a polynucleotide comprising a polynucleotide sequence selected from the group consisting of SEQ ID NO:13-24, ii) a polynucleotide comprising a naturally occurring polynucleotide sequence at least 90% identical to a polynucleotide sequence selected from the group consisting of SEQ ID NO:13-24, iii) a polynucleotide complementary to the polynucleotide of i), iv) a polynucleotide complementary to the polynucleotide of ii), and v) an RNA equivalent of i)-iv). Alternatively, the target polynucleotide comprises a fragment of a polynucleotide sequence selected from the group consisting of i)-v) above; c) quantifying the amount of hybridization complex; and d) comparing the amount of hybridization complex in the treated biological sample with the amount of hybridization complex in an untreated biological sample, wherein a difference in the amount of hybridization complex in the treated biological sample is indicative of toxicity of the test compound.

BRIEF DESCRIPTION OF THE TABLES

Table 1 summarizes the nomenclature for the full length polynucleotide and polypeptide sequences of the present invention.

Table 2 shows the GenBank identification number and annotation of the nearest GenBank homolog for polypeptides of the invention. The probability score for the match between each polypeptide and its GenBank homolog is also shown.

Table 3 shows structural features of polypeptide sequences of the invention, including predicted motifs and domains, along with the methods, algorithms, and searchable databases used for analysis of the polypeptides.

Table 4 lists the cDNA and/or genomic DNA fragments which were used to assemble polynucleotide sequences of the invention, along with selected fragments of the polynucleotide sequences.

Table 5 shows the representative cDNA library for polynucleotides of the invention.

Table 6 provides an appendix which describes the tissues and vectors used for construction of the cDNA libraries shown in Table 5.

Table 7 shows the tools, programs, and algorithms used to analyze the polynucleotides and polypeptides of the invention, along with applicable descriptions, references, and threshold parameters.

DESCRIPTION OF THE INVENTION

Before the present proteins, nucleotide sequences, and methods are described, it is understood that this invention is not limited to the particular machines, materials and methods described, as these may vary. It is also to be understood that the terminology used herein is for the purpose of describing particular embodiments only, and is not intended to limit the scope of the present invention which will be limited only by the appended claims.

It must be noted that as used herein and in the appended claims, the singular forms "a," "an," and "the" include plural reference unless the context clearly dictates otherwise. Thus, for example, a reference to "a host cell" includes a plurality of such host cells, and a reference to "an antibody" is a reference to one or more antibodies and equivalents thereof known to those skilled in the art, and so forth.

Unless defined otherwise, all technical and scientific terms used herein have the same meanings as commonly understood by one of ordinary skill in the art to which this invention belongs. Although any machines, materials, and methods similar or equivalent to those described herein can be used to practice or test the present invention, the preferred machines, materials and methods are now described. All publications mentioned herein are cited for the purpose of describing and disclosing the cell lines, protocols, reagents and vectors which are reported in the publications and which might be used in connection with the invention. Nothing herein is to be construed as an admission that the invention is not entitled to antedate such disclosure by virtue of prior invention.

DEFINITIONS

"PP" refers to the amino acid sequences of substantially purified PP obtained from any species, particularly a mammalian species, including bovine, ovine, porcine, murine, equine, and human, and from any source, whether natural, synthetic, semi-synthetic, or recombinant.

The term "agonist" refers to a molecule which intensifies or mimics the biological activity of PP. Agonists may include proteins, nucleic acids, carbohydrates, small molecules, or any other compound or composition which modulates the activity of PP either by directly interacting with PP or by acting on components of the biological pathway in which PP participates.

An "allelic variant" is an alternative form of the gene encoding PP. Allelic variants may result from at least one mutation in the nucleic acid sequence and may result in altered mRNAs or in polypeptides whose structure or function may or may not be altered. A gene may have none, one, or many allelic variants of its naturally occurring form. Common mutational changes which give rise to allelic variants are generally ascribed to natural deletions, additions, or substitutions of nucleotides. Each of these types of changes may occur alone, or in combination with the others, one or more times in a given sequence.

"Altered" nucleic acid sequences encoding PP include those sequences with deletions, insertions, or substitutions of different nucleotides, resulting in a polypeptide the same as PP or a polypeptide with at least one functional characteristic of PP. Included within this definition are polymorphisms which may or may not be readily detectable using a particular oligonucleotide probe of the polynucleotide encoding PP, and improper or unexpected hybridization to allelic variants, with a locus other than the normal chromosomal locus for the polynucleotide sequence encoding PP. The encoded protein may also be "altered," and may contain deletions, insertions, or substitutions of amino acid residues which produce a silent change and result in a functionally equivalent PP. Deliberate amino acid substitutions may be made on the basis of similarity in polarity, charge, solubility, hydrophobicity, hydrophilicity, and/or the amphipathic nature of the residues, as long as the biological or immunological activity of PP is retained. For example, negatively charged amino acids may include aspartic acid and glutamic acid, and positively charged amino acids may include lysine and arginine. Amino acids with uncharged polar side chains having similar hydrophilicity values may include: asparagine and glutamine; and serine and threonine. Amino acids with uncharged side chains having similar hydrophilicity values may include: leucine, isoleucine, and valine; glycine and alanine; and phenylalanine and tyrosine.

The terms "amino acid" and "amino acid sequence" refer to an oligopeptide, peptide, polypeptide, or protein sequence, or a fragment of any of these, and to naturally occurring or synthetic molecules. Where "amino acid sequence" is recited to refer to a sequence of a naturally occurring protein molecule, "amino acid sequence" and like terms are not meant to limit the amino acid sequence to the complete native amino acid sequence associated with the recited protein molecule.

"Amplification" relates to the production of additional copies of a nucleic acid sequence. Amplification is generally carried out using polymerase chain reaction (PCR) technologies well known in the art.

The term "antagonist" refers to a molecule which inhibits or attenuates the biological activity of PP. Antagonists may include proteins such as antibodies, nucleic acids, carbohydrates, small molecules, or any other compound or composition which modulates the activity of PP either by directly interacting with PP or by acting on components of the biological pathway in which PP participates.

The term "antibody" refers to intact immunoglobulin molecules as well as to fragments thereof, such as Fab, $F(ab')_2$, and Fv fragments, which are capable of binding an epitopic determinant. Antibodies that bind PP polypeptides can be prepared using intact polypeptides or using fragments containing small peptides of interest as the immunizing antigen. The polypeptide or oligopeptide used to immunize an animal (e.g., a mouse, a rat, or a rabbit) can be derived from the translation of RNA, or synthesized chemically, and can be conjugated to a carrier protein if desired. Commonly used carriers that are chemically coupled to peptides include bovine serum albumin, thyroglobulin, and keyhole limpet hemocyanin (KLH). The coupled peptide is then used to immunize the animal.

The term "antigenic determinant" refers to that region of a molecule (i.e., an epitope) that makes contact with a particular antibody. When a protein or a fragment of a protein is used to immunize a host animal, numerous regions of the protein may induce the production of antibodies which bind specifically to antigenic determinants (particular regions or three-dimensional structures on the protein). An antigenic determinant may compete with the intact antigen (i.e., the immunogen used to elicit the immune response) for binding to an antibody.

The term "aptamer" refers to a nucleic acid or oligonucleotide molecule that binds to a specific molecular target. Aptamers are derived from an in vitro evolutionary process (e.g., SELEX (Systematic Evolution of Ligands by EXponential Enrichment), described in U.S. Patent No. 5,270,163), which selects for target-specific aptamer sequences from large combinatorial libraries. Aptamer compositions may be double-stranded or single-stranded, and may include deoxyribonucleotides, ribonucleotides, nucleotide derivatives, or other nucleotide-like molecules. The nucleotide components of an aptamer may have modified sugar groups (e.g., the 2'-OH group of a ribonucleotide may be replaced by 2'-F or 2'-NH_2), which may improve a desired property, e.g., resistance to nucleases or longer lifetime in blood. Aptamers may be conjugated to other molecules, e.g., a high molecular weight carrier to slow clearance of the aptamer from the circulatory system. Aptamers may be specifically cross-linked to their cognate ligands, e.g., by photo-activation of a cross-linker. (See, e.g., Brody, E.N. and L. Gold (2000) J. Biotechnol. 74:5-13.)

The term "intramer" refers to an aptamer which is expressed in vivo. For example, a vaccinia virus-based RNA expression system has been used to express specific RNA aptamers at high levels in the cytoplasm of leukocytes (Blind, M. et al. (1999) Proc. Natl Acad. Sci. USA 96:3606-3610).

The term "spiegelmer" refers to an aptamer which includes L-DNA, L-RNA, or other left-handed nucleotide derivatives or nucleotide-like molecules. Aptamers containing left-handed nucleotides are resistant to degradation by naturally occurring enzymes, which normally act on substrates containing right-handed nucleotides.

The term "antisense" refers to any composition capable of base-pairing with the "sense" (coding) strand of a specific nucleic acid sequence. Antisense compositions may include DNA; RNA; peptide nucleic acid (PNA); oligonucleotides having modified backbone linkages such as phosphorothioates, methylphosphonates, or benzylphosphonates; oligonucleotides having modified sugar groups such as 2'-methoxyethyl sugars or 2'-methoxyethoxy sugars; or oligonucleotides having modified bases such as 5-methyl cytosine, 2'-deoxyuracil, or 7-deaza-2'-deoxyguanosine. Antisense molecules may be produced by any method including chemical synthesis or transcription. Once introduced into a cell, the complementary antisense molecule base-pairs with a naturally occurring nucleic acid sequence produced by the cell to form duplexes which block either transcription or translation. The designation "negative" or "minus" can refer to the antisense strand, and the designation "positive" or "plus" can refer to the sense strand of a reference DNA molecule.

The term "biologically active" refers to a protein having structural, regulatory, or biochemical functions of a naturally occurring molecule. Likewise, "immunologically active" or "immunogenic" refers to the capability of the natural, recombinant, or synthetic PP, or of any oligopeptide thereof, to induce a specific immune response in appropriate animals or cells and to bind with specific antibodies.

"Complementary" describes the relationship between two single-stranded nucleic acid sequences that anneal by base-pairing. For example, 5'-AGT-3' pairs with its complement, 3'-TCA-5'.

A "composition comprising a given polynucleotide sequence" and a "composition comprising a given amino acid sequence" refer broadly to any composition containing the given polynucleotide or amino acid sequence. The composition may comprise a dry formulation or an aqueous solution. Compositions comprising polynucleotide sequences encoding PP or fragments of PP may be employed as hybridization probes. The probes may be stored in freeze-dried form and may be associated with a stabilizing agent such as a carbohydrate. In hybridizations, the probe may be deployed in an aqueous solution containing salts (e.g., NaCl), detergents (e.g., sodium dodecyl sulfate; SDS), and other components (e.g., Denhardt's solution, dry milk, salmon sperm DNA, etc.).

"Consensus sequence" refers to a nucleic acid sequence which has been subjected to repeated DNA sequence analysis to resolve uncalled bases, extended using the XL-PCR kit (Applied Biosystems, Foster City CA) in the 5' and/or the 3' direction, and resequenced, or which has been assembled from one or more overlapping cDNA, EST, or genomic DNA fragments using a computer program for fragment assembly, such as the GELVIEW fragment assembly system (GCG, Madison WI) or Phrap (University of Washington, Seattle WA). Some sequences have been both extended and assembled to produce the consensus sequence.

"Conservative amino acid substitutions" are those substitutions that are predicted to least interfere with the properties of the original protein, i.e., the structure and especially the function of the protein is conserved and not significantly changed by such substitutions. The table below shows amino acids which may be substituted for an original amino acid in a protein and which are regarded as conservative amino acid substitutions.

Original Residue	Conservative Substitution
Ala	Gly, Ser
Arg	His, Lys
Asn	Asp, Gln, His
Asp	Asn, Glu

Cys	Ala, Ser
Gln	Asn, Glu, His
Glu	Asp, Gln, His
Gly	Ala
His	Asn, Arg, Gln, Glu
Ile	Leu, Val
Leu	Ile, Val
Lys	Arg, Gln, Glu
Met	Leu, Ile
Phe	His, Met, Leu, Trp, Tyr
Ser	Cys, Thr
Thr	Ser, Val
Trp	Phe, Tyr
Tyr	His, Phe, Trp
Val	Ile, Leu, Thr

Conservative amino acid substitutions generally maintain (a) the structure of the polypeptide backbone in the area of the substitution, for example, as a beta sheet or alpha helical conformation, (b) the charge or hydrophobicity of the molecule at the site of the substitution, and/or (c) the bulk of the side chain.

A "deletion" refers to a change in the amino acid or nucleotide sequence that results in the absence of one or more amino acid residues or nucleotides.

The term "derivative" refers to a chemically modified polynucleotide or polypeptide. Chemical modifications of a polynucleotide can include, for example, replacement of hydrogen by an alkyl, acyl, hydroxyl, or amino group. A derivative polynucleotide encodes a polypeptide which retains at least one biological or immunological function of the natural molecule. A derivative polypeptide is one modified by glycosylation, pegylation, or any similar process that retains at least one biological or immunological function of the polypeptide from which it was derived.

A "detectable label" refers to a reporter molecule or enzyme that is capable of generating a measurable signal and is covalently or noncovalently joined to a polynucleotide or polypeptide.

"Differential expression" refers to increased or upregulated; or decreased, downregulated, or absent gene or protein expression, determined by comparing at least two different samples. Such comparisons may be carried out between, for example, a treated and an untreated sample, or a diseased and a normal sample.

"Exon shuffling" refers to the recombination of different coding regions (exons). Since an exon may represent a structural or functional domain of the encoded protein, new proteins may be assembled through the novel reassortment of stable substructures, thus allowing acceleration of the evolution of new protein functions.

A "fragment" is a unique portion of PP or the polynucleotide encoding PP which is identical in sequence to but shorter in length than the parent sequence. A fragment may comprise up to the entire length of the defined sequence, minus one nucleotide/amino acid residue. For example, a fragment may comprise from 5 to 1000 contiguous nucleotides or amino acid residues. A fragment used as a probe, primer, antigen, therapeutic molecule, or for other purposes, may be at least 5, 10, 15, 16, 20, 25, 30, 40, 50, 60, 75, 100, 150, 250 or at least 500 contiguous nucleotides or amino acid residues in length. Fragments may be preferentially selected from certain regions of a molecule. For example, a polypeptide fragment may comprise a certain length of contiguous amino acids selected from the first 250 or 500 amino acids (or first 25% or 50%) of a polypeptide as shown in a certain defined sequence. Clearly these lengths are exemplary, and any length that is supported by the specification, including the Sequence Listing, tables, and figures, may be encompassed by the present embodiments.

A fragment of SEQ ID NO:13-24 comprises a region of unique polynucleotide sequence that specifically identifies SEQ ID NO:13-24, for example, as distinct from any other sequence in the genome from which the fragment was obtained. A fragment of SEQ ID NO:13-24 is useful, for example, in hybridization and amplification technologies and in analogous methods that distinguish SEQ ID NO:13-24 from related polynucleotide sequences. The precise length of a fragment of SEQ ID NO:13-24 and the region of SEQ ID NO:13-24 to which the fragment corresponds are routinely determinable by one of ordinary skill in the art based on the intended purpose for the fragment.

A fragment of SEQ ID NO:1-12 is encoded by a fragment of SEQ ID NO:13-24. A fragment of SEQ ID NO:1-12 comprises a region of unique amino acid sequence that specifically identifies SEQ ID NO:1-12. For example, a fragment of SEQ ID NO:1-12 is useful as an immunogenic peptide for the development of antibodies that specifically recognize SEQ ID NO:1-12. The precise length of a fragment of SEQ ID NO:1-12 and the region of SEQ ID NO:1-12 to which the fragment corresponds are routinely determinable by one of ordinary skill in the art based on the intended purpose for the fragment.

A "full length" polynucleotide sequence is one containing at least a translation initiation codon (e.g., methionine) followed by an open reading frame and a translation termination codon. A "full length" polynucleotide sequence encodes a "full length" polypeptide sequence.

"Homology" refers to sequence similarity or, interchangeably, sequence identity, between two or more polynucleotide sequences or two or more polypeptide sequences.

The terms "percent identity" and "% identity," as applied to polynucleotide sequences, refer to the percentage of residue matches between at least two polynucleotide sequences aligned using a standardized algorithm. Such an algorithm may insert, in a standardized and reproducible way, gaps in the sequences being compared in order to optimize alignment between two sequences, and therefore achieve a more meaningful comparison of the two sequences.

Percent identity between polynucleotide sequences may be determined using the default parameters of the CLUSTAL V algorithm as incorporated into the MEGALIGN version 3.12e sequence alignment program. This program is part of the LASERGENE software package, a suite of molecular biological analysis programs (DNASTAR, Madison WI). CLUSTAL V is described in Higgins, D.G. and P.M. Sharp (1989) CABIOS 5:151-153 and in Higgins, D.G. et al. (1992) CABIOS 8:189-191. For pairwise alignments of polynucleotide sequences, the default parameters are set as follows: Ktuple=2, gap penalty=5, window=4, and "diagonals saved"=4. The "weighted" residue weight table is selected as the default. Percent identity is reported by CLUSTAL V as the "percent similarity" between aligned polynucleotide sequences.

Alternatively, a suite of commonly used and freely available sequence comparison algorithms is provided by the National Center for Biotechnology Information (NCBI) Basic Local Alignment Search Tool (BLAST) (Altschul, S.F. et al. (1990) J. Mol. Biol. 215:403-410), which is available from several sources, including the NCBI, Bethesda, MD, and on the Internet at http://www.ncbi.nlm.nih.gov/BLAST/. The BLAST software suite includes various sequence analysis programs including "blastn," that is used to align a known polynucleotide sequence with other polynucleotide sequences from a variety of databases. Also available is a tool called "BLAST 2 Sequences" that is used for direct pairwise comparison of two nucleotide sequences. "BLAST 2 Sequences" can be accessed and used interactively at http://www.ncbi.nlm.nih.gov/gorf/bl2.html. The "BLAST 2 Sequences" tool can be used for both blastn and blastp (discussed below). BLAST programs are commonly used with gap and other parameters set to default settings. For example, to compare two nucleotide sequences, one may use blastn with the "BLAST 2 Sequences" tool Version 2.0.12 (April-21-2000) set at default parameters. Such default parameters may be, for example:

Matrix: BLOSUM62
Reward for match: 1
Penalty for mismatch: -2
Open Gap: 5 and Extension Gap: 2 penalties
Gap x drop-off: 50
Expect: 10
Word Size: 11
Filter: on

Percent identity may be measured over the length of an entire defined sequence, for example, as defined by a particular SEQ ID number, or may be measured over a shorter length, for example, over the length of a fragment taken from a larger, defined sequence, for instance, a fragment of at least 20, at least 30, at least 40, at least 50, at least 70, at least 100, or at least 200 contiguous nucleotides. Such lengths are exemplary only, and it is understood that any fragment length supported by the sequences shown herein, in the tables, figures, or Sequence Listing, may be used to describe a length over which percentage identity may be measured.

Nucleic acid sequences that do not show a high degree of identity may nevertheless encode similar amino acid sequences due to the degeneracy of the genetic code. It is understood that changes in a nucleic acid sequence can be made using this degeneracy to produce multiple nucleic acid sequences that all encode substantially the same protein.

The phrases "percent identity" and "% identity," as applied to polypeptide sequences, refer to the percentage of residue matches between at least two polypeptide sequences aligned using a standardized algorithm. Methods of polypeptide sequence alignment are well-known. Some alignment methods take into account conservative amino acid substitutions. Such conservative substitutions, explained in more detail above, generally preserve the charge and hydrophobicity at the site of substitution, thus preserving the structure (and therefore function) of the polypeptide.

Percent identity between polypeptide sequences may be determined using the default parameters of the CLUSTAL V algorithm as incorporated into the MEGALIGN version 3.12e sequence alignment program (described and referenced above). For pairwise alignments of polypeptide sequences using CLUSTAL V, the default parameters are set as follows: Ktuple=1, gap penalty=3, window=5, and "diagonals saved"=5. The PAM250 matrix is selected as the default residue weight table. As with polynucleotide alignments, the percent identity is reported by CLUSTAL V as the "percent similarity" between aligned polypeptide sequence pairs.

Alternatively the NCBI BLAST software suite may be used. For example, for a pairwise comparison of two polypeptide sequences, one may use the "BLAST 2 Sequences" tool Version 2.0.12 (April-21-2000) with blastp set at default parameters. Such default parameters may be, for example:

Matrix: BLOSUM62

Open Gap: 11 and Extension Gap: 1 penalties

Gap x drop-off: 50

Expect: 10

Word Size: 3

Filter: on

Percent identity may be measured over the length of an entire defined polypeptide sequence, for example, as defined by a particular SEQ ID number, or may be measured over a shorter length, for example, over the length of a fragment taken from a larger, defined polypeptide sequence, for instance, a fragment of at least 15, at least 20, at least 30, at least 40, at least 50, at least 70 or at least 150 contiguous residues. Such lengths are exemplary only, and it is understood that any fragment length supported by the sequences shown herein, in the tables, figures or Sequence Listing, may be used to describe a length over which percentage identity may be measured.

"Human artificial chromosomes" (HACs) are linear microchromosomes which may contain DNA sequences of about 6 kb to 10 Mb in size and which contain all of the elements required for chromosome replication, segregation and maintenance.

The term "humanized antibody" refers to an antibody molecule in which the amino acid sequence in the non-antigen binding regions has been altered so that the antibody more closely resembles a human antibody, and still retains its original binding ability.

"Hybridization" refers to the process by which a polynucleotide strand anneals with a complementary strand through base pairing under defined hybridization conditions. Specific hybridization is an indication that two nucleic acid sequences share a high degree of complementarity. Specific hybridization complexes form under permissive annealing conditions and remain hybridized after the "washing" step(s). The washing step(s) is particularly important in determining the stringency of the hybridization process, with more stringent conditions allowing less non-specific binding, i.e., binding between pairs of nucleic acid strands that are not perfectly matched. Permissive conditions for annealing of nucleic acid sequences are routinely determinable by one of ordinary skill in the art and may be consistent among hybridization experiments, whereas wash conditions may be varied among experiments to achieve the desired stringency, and therefore hybridization specificity. Permissive annealing conditions occur, for example, at 68°C in the presence of about 6 x SSC, about 1% (w/v) SDS, and about 100 μg/ml sheared, denatured salmon sperm DNA.

Generally, stringency of hybridization is expressed, in part, with reference to the temperature under which the wash step is carried out. Such wash temperatures are typically selected to be about 5°C to 20°C lower than the thermal melting point (T_m) for the specific sequence at a defined ionic strength and pH. The T_m is the temperature (under defined ionic strength and pH) at which 50% of the target sequence hybridizes to a perfectly matched probe. An equation for calculating T_m and conditions for nucleic acid hybridization are well known and can be found in Sambrook, J. et al. (1989) Molecular Cloning: A Laboratory Manual, 2nd ed., vol. 1-3, Cold Spring Harbor Press, Plainview NY; specifically see volume 2, chapter 9.

High stringency conditions for hybridization between polynucleotides of the present invention include wash conditions of 68°C in the presence of about 0.2 x SSC and about 0.1% SDS, for 1 hour. Alternatively, temperatures of about 65°C, 60°C, 55°C, or 42°C may be used. SSC concentration may be varied from about 0.1 to 2 x SSC, with SDS being present at about 0.1%. Typically, blocking reagents are used to block non-specific hybridization. Such blocking reagents include, for instance, sheared and denatured salmon sperm DNA at about 100-200 μg/ml. Organic solvent, such as formamide at a concentration of about 35-50% v/v, may also be used under particular circumstances, such as for RNA:DNA hybridizations. Useful variations on these wash conditions will be readily apparent to those of ordinary skill in the art. Hybridization, particularly under high stringency conditions, may be suggestive of evolutionary similarity between the nucleotides. Such similarity is strongly indicative of a similar role for the nucleotides and their encoded polypeptides.

The term "hybridization complex" refers to a complex formed between two nucleic acid sequences by virtue of the formation of hydrogen bonds between complementary bases. A hybridization complex may be formed in solution (e.g., C_0t or R_0t analysis) or formed between one nucleic acid sequence present in solution and another nucleic acid sequence immobilized on a solid support (e.g., paper, membranes, filters, chips, pins or glass slides, or any other appropriate substrate to which cells or their nucleic acids have been fixed).

The words "insertion" and "addition" refer to changes in an amino acid or nucleotide sequence resulting in the addition of one or more amino acid residues or nucleotides, respectively.

"Immune response" can refer to conditions associated with inflammation, trauma, immune disorders, or infectious or genetic disease, etc. These conditions can be characterized by expression of various factors, e.g., cytokines, chemokines, and other signaling molecules, which may affect cellular and systemic defense systems.

An "immunogenic fragment" is a polypeptide or oligopeptide fragment of PP which is capable of eliciting an immune response when introduced into a living organism, for example, a mammal. The term "immunogenic fragment" also includes any polypeptide or oligopeptide fragment of PP which is useful in any of the antibody production methods disclosed herein or known in the art.

The term "microarray" refers to an arrangement of a plurality of polynucleotides, polypeptides, or other chemical compounds on a substrate.

The terms "element" and "array element" refer to a polynucleotide, polypeptide, or other chemical compound having a unique and defined position on a microarray.

The term "modulate" refers to a change in the activity of PP. For example, modulation may cause an increase or a decrease in protein activity, binding characteristics, or any other biological, functional, or immunological properties of PP.

The phrases "nucleic acid" and "nucleic acid sequence" refer to a nucleotide, oligonucleotide, polynucleotide, or any fragment thereof. These phrases also refer to DNA or RNA of genomic or synthetic origin which may be single-stranded or double-stranded and may represent the sense or the antisense strand, to peptide nucleic acid (PNA), or to any DNA-like or RNA-like material.

"Operably linked" refers to the situation in which a first nucleic acid sequence is placed in a functional relationship with a second nucleic acid sequence. For instance, a promoter is operably linked to a coding sequence if the promoter affects the transcription or expression of the coding sequence. Operably linked DNA sequences may be in close proximity or contiguous and, where necessary to join two protein coding regions, in the same reading frame.

"Peptide nucleic acid" (PNA) refers to an antisense molecule or anti-gene agent which comprises an oligonucleotide of at least about 5 nucleotides in length linked to a peptide backbone of amino acid residues ending in lysine. The terminal lysine confers solubility to the composition. PNAs preferentially bind complementary single stranded DNA or RNA and stop transcript elongation, and may be pegylated to extend their lifespan in the cell.

"Post-translational modification" of an PP may involve lipidation, glycosylation, phosphorylation, acetylation, racemization, proteolytic cleavage, and other modifications known in the art. These processes may occur synthetically or biochemically. Biochemical modifications will vary by cell type depending on the enzymatic milieu of PP.

"Probe" refers to nucleic acid sequences encoding PP, their complements, or fragments thereof, which are used to detect identical, allelic or related nucleic acid sequences. Probes are isolated oligonucleotides or polynucleotides attached to a detectable label or reporter molecule. Typical labels include radioactive isotopes, ligands, chemiluminescent agents, and enzymes. "Primers" are short nucleic acids, usually DNA oligonucleotides, which may be annealed to a target polynucleotide by complementary base-pairing. The primer may then be extended along the target DNA strand by a DNA polymerase enzyme. Primer pairs can be used for amplification (and identification) of a nucleic acid sequence, e.g., by the polymerase chain reaction (PCR).

Probes and primers as used in the present invention typically comprise at least 15 contiguous nucleotides of a known sequence. In order to enhance specificity, longer probes and primers may also be employed, such as probes and primers that comprise at least 20, 25, 30, 40, 50, 60, 70, 80, 90, 100, or at least 150 consecutive nucleotides of the disclosed nucleic acid sequences. Probes and primers may be considerably longer than these examples, and it is understood that any length supported by the specification, including the tables, figures, and Sequence Listing, may be used.

Methods for preparing and using probes and primers are described in the references, for example Sambrook, J. et al. (1989) Molecular Cloning: A Laboratory Manual, 2nd ed., vol. 1-3, Cold Spring Harbor Press, Plainview NY; Ausubel, F.M. et al. (1987) Current Protocols in Molecular Biology, Greene Publ. Assoc. & Wiley-Intersciences, New York NY; Innis, M. et al. (1990) PCR Protocols, A Guide to Methods and Applications, Academic Press, San Diego CA. PCR primer pairs can be derived from a known sequence, for example, by using computer programs intended for that purpose such as Primer (Version 0.5, 1991, Whitehead Institute for Biomedical Research, Cambridge MA).

Oligonucleotides for use as primers are selected using software known in the art for such purpose. For example, OLIGO 4.06 software is useful for the selection of PCR primer pairs of up to

100 nucleotides each, and for the analysis of oligonucleotides and larger polynucleotides of up to 5,000 nucleotides from an input polynucleotide sequence of up to 32 kilobases. Similar primer selection programs have incorporated additional features for expanded capabilities. For example, the PrimOU primer selection program (available to the public from the Genome Center at University of Texas South West Medical Center, Dallas TX) is capable of choosing specific primers from megabase sequences and is thus useful for designing primers on a genome-wide scope. The Primer3 primer selection program (available to the public from the Whitehead Institute/MIT Center for Genome Research, Cambridge MA) allows the user to input a "mispriming library," in which sequences to avoid as primer binding sites are user-specified. Primer3 is useful, in particular, for the selection of oligonucleotides for microarrays. (The source code for the latter two primer selection programs may also be obtained from their respective sources and modified to meet the user's specific needs.) The PrimeGen program (available to the public from the UK Human Genome Mapping Project Resource Centre, Cambridge UK) designs primers based on multiple sequence alignments, thereby allowing selection of primers that hybridize to either the most conserved or least conserved regions of aligned nucleic acid sequences. Hence, this program is useful for identification of both unique and conserved oligonucleotides and polynucleotide fragments. The oligonucleotides and polynucleotide fragments identified by any of the above selection methods are useful in hybridization technologies, for example, as PCR or sequencing primers, microarray elements, or specific probes to identify fully or partially complementary polynucleotides in a sample of nucleic acids. Methods of oligonucleotide selection are not limited to those described above.

A "recombinant nucleic acid" is a sequence that is not naturally occurring or has a sequence that is made by an artificial combination of two or more otherwise separated segments of sequence. This artificial combination is often accomplished by chemical synthesis or, more commonly, by the artificial manipulation of isolated segments of nucleic acids, e.g., by genetic engineering techniques such as those described in Sambrook, supra. The term recombinant includes nucleic acids that have been altered solely by addition, substitution, or deletion of a portion of the nucleic acid. Frequently, a recombinant nucleic acid may include a nucleic acid sequence operably linked to a promoter sequence. Such a recombinant nucleic acid may be part of a vector that is used, for example, to transform a cell.

Alternatively, such recombinant nucleic acids may be part of a viral vector, e.g., based on a vaccinia virus, that could be use to vaccinate a mammal wherein the recombinant nucleic acid is expressed, inducing a protective immunological response in the mammal.

A "regulatory element" refers to a nucleic acid sequence usually derived from untranslated regions of a gene and includes enhancers, promoters, introns, and 5' and 3' untranslated regions

(UTRs). Regulatory elements interact with host or viral proteins which control transcription, translation, or RNA stability.

"Reporter molecules" are chemical or biochemical moieties used for labeling a nucleic acid, amino acid, or antibody. Reporter molecules include radionuclides; enzymes; fluorescent, chemiluminescent, or chromogenic agents; substrates; cofactors; inhibitors; magnetic particles; and other moieties known in the art.

An "RNA equivalent," in reference to a DNA sequence, is composed of the same linear sequence of nucleotides as the reference DNA sequence with the exception that all occurrences of the nitrogenous base thymine are replaced with uracil, and the sugar backbone is composed of ribose instead of deoxyribose.

The term "sample" is used in its broadest sense. A sample suspected of containing PP, nucleic acids encoding PP, or fragments thereof may comprise a bodily fluid; an extract from a cell, chromosome, organelle, or membrane isolated from a cell; a cell; genomic DNA, RNA, or cDNA, in solution or bound to a substrate; a tissue; a tissue print; etc.

The terms "specific binding" and "specifically binding" refer to that interaction between a protein or peptide and an agonist, an antibody, an antagonist, a small molecule, or any natural or synthetic binding composition. The interaction is dependent upon the presence of a particular structure of the protein, e.g., the antigenic determinant or epitope, recognized by the binding molecule. For example, if an antibody is specific for epitope "A," the presence of a polypeptide comprising the epitope A, or the presence of free unlabeled A, in a reaction containing free labeled A and the antibody will reduce the amount of labeled A that binds to the antibody.

The term "substantially purified" refers to nucleic acid or amino acid sequences that are removed from their natural environment and are isolated or separated, and are at least 60% free, preferably at least 75% free, and most preferably at least 90% free from other components with which they are naturally associated.

A "substitution" refers to the replacement of one or more amino acid residues or nucleotides by different amino acid residues or nucleotides, respectively.

"Substrate" refers to any suitable rigid or semi-rigid support including membranes, filters, chips, slides, wafers, fibers, magnetic or nonmagnetic beads, gels, tubing, plates, polymers, microparticles and capillaries. The substrate can have a variety of surface forms, such as wells, trenches, pins, channels and pores, to which polynucleotides or polypeptides are bound.

A "transcript image" refers to the collective pattern of gene expression by a particular cell type or tissue under given conditions at a given time.

"Transformation" describes a process by which exogenous DNA is introduced into a recipient cell. Transformation may occur under natural or artificial conditions according to various methods well known in the art, and may rely on any known method for the insertion of foreign nucleic acid sequences into a prokaryotic or eukaryotic host cell. The method for transformation is selected based on the type of host cell being transformed and may include, but is not limited to, bacteriophage or viral infection, electroporation, heat shock, lipofection, and particle bombardment. The term "transformed cells" includes stably transformed cells in which the inserted DNA is capable of replication either as an autonomously replicating plasmid or as part of the host chromosome, as well as transiently transformed cells which express the inserted DNA or RNA for limited periods of time.

A "transgenic organism," as used herein, is any organism, including but not limited to animals and plants, in which one or more of the cells of the organism contains heterologous nucleic acid introduced by way of human intervention, such as by transgenic techniques well known in the art. The nucleic acid is introduced into the cell, directly or indirectly by introduction into a precursor of the cell, by way of deliberate genetic manipulation, such as by microinjection or by infection with a recombinant virus. The term genetic manipulation does not include classical cross-breeding, or in vitro fertilization, but rather is directed to the introduction of a recombinant DNA molecule. The transgenic organisms contemplated in accordance with the present invention include bacteria, cyanobacteria, fungi, plants and animals. The isolated DNA of the present invention can be introduced into the host by methods known in the art, for example infection, transfection, transformation or transconjugation. Techniques for transferring the DNA of the present invention into such organisms are widely known and provided in references such as Sambrook et al. (1989), supra.

A "variant" of a particular nucleic acid sequence is defined as a nucleic acid sequence having at least 40% sequence identity to the particular nucleic acid sequence over a certain length of one of the nucleic acid sequences using blastn with the "BLAST 2 Sequences" tool Version 2.0.9 (May-07-1999) set at default parameters. Such a pair of nucleic acids may show, for example, at least 50%, at least 60%, at least 70%, at least 80%, at least 85%, at least 90%, at least 91%, at least 92%, at least 93%, at least 94%, at least 95%, at least 96%, at least 97%, at least 98%, or at least 99% or greater sequence identity over a certain defined length. A variant may be described as, for example, an "allelic" (as defined above), "splice," "species," or "polymorphic" variant. A splice variant may have significant identity to a reference molecule, but will generally have a greater or lesser number of polynucleotides due to alternate splicing of exons during mRNA processing. The corresponding polypeptide may possess additional functional domains or lack domains that are present in the reference molecule. Species variants are polynucleotide sequences that vary from one species to another. The resulting polypeptides will generally have significant amino acid identity relative to each other. A polymorphic variant is a variation in the polynucleotide sequence of a particular gene between individuals of a given species. Polymorphic variants also may encompass "single nucleotide polymorphisms" (SNPs) in which the polynucleotide sequence varies by one nucleotide base. The presence of SNPs may be indicative of, for example, a certain population, a disease state, or a propensity for a disease state.

A "variant" of a particular polypeptide sequence is defined as a polypeptide sequence having at least 40% sequence identity to the particular polypeptide sequence over a certain length of one of the polypeptide sequences using blastp with the "BLAST 2 Sequences" tool Version 2.0.9 (May-07-1999) set at default parameters. Such a pair of polypeptides may show, for example, at least 50%, at least 60%, at least 70%, at least 80%, at least 90%, at least 91%, at least 92%, at least 93%, at least 94%, at least 95%, at least 96%, at least 97%, at least 98%, or at least 99% or greater sequence identity over a certain defined length of one of the polypeptides.

THE INVENTION

The invention is based on the discovery of new human protein phosphatases (PP), the polynucleotides encoding PP, and the use of these compositions for the diagnosis, treatment, or prevention of immune system disorders, neurological disorders, developmental disorders, and cell proliferative disorders, including cancer.

Table 1 summarizes the nomenclature for the full length polynucleotide and polypeptide sequences of the invention. Each polynucleotide and its corresponding polypeptide are correlated to a single Incyte project identification number (Incyte Project ID). Each polypeptide sequence is denoted by both a polypeptide sequence identification number (Polypeptide SEQ ID NO:) and an Incyte polypeptide sequence number (Incyte Polypeptide ID) as shown. Each polynucleotide sequence is denoted by both a polynucleotide sequence identification number (Polynucleotide SEQ ID NO:) and an Incyte polynucleotide consensus sequence number (Incyte Polynucleotide ID) as shown.

Table 2 shows sequences with homology to the polypeptides of the invention as identified by BLAST analysis against the GenBank protein (genpept) database. Columns 1 and 2 show the polypeptide sequence identification number (Polypeptide SEQ ID NO:) and the corresponding Incyte polypeptide sequence number (Incyte Polypeptide ID) for polypeptides of the invention. Column 3 shows the GenBank identification number (Genbank ID NO:) of the nearest GenBank homolog. Column 4 shows the probability score for the match between each polypeptide and its GenBank homolog. Column 5 shows the annotation of the GenBank homolog along with relevant citations where applicable, all of which are expressly incorporated by reference herein.

Table 3 shows various structural features of the polypeptides of the invention. Columns 1 and 2 show the polypeptide sequence identification number (SEQ ID NO:) and the corresponding Incyte polypeptide sequence number (Incyte Polypeptide ID) for each polypeptide of the invention. Column 3 shows the number of amino acid residues in each polypeptide. Column 4 shows potential phosphorylation sites, and column 5 shows potential glycosylation sites, as determined by the MOTIFS program of the GCG sequence analysis software package (Genetics Computer Group, Madison WI). Column 6 shows amino acid residues comprising signature sequences, domains, and motifs. Column 7 shows analytical methods for protein structure/function analysis and in some cases, searchable databases to which the analytical methods were applied.

Together, Tables 2 and 3 summarize the properties of polypeptides of the invention, and these properties establish that the claimed polypeptides are protein phosphatases. For example, SEQ ID NO:2 is 47% identical to *Escherichia coli* Serine/Threonine protein phosphatase (EC 3.1.3.16) (GenBank ID g1736483) as determined by the Basic Local Alignment Search Tool (BLAST). (See Table 2.) The BLAST probability score is 8.4e-49, which indicates the probability of obtaining the observed polypeptide sequence alignment by chance. SEQ ID NO:2 also contains a serine/threonine specific protein phosphatases signature as indicated in the PROFILESCAN analysis. (See Table 3.) Data from MOTIFS analysis provides further corroborative evidence that SEQ ID NO:2 is a serine/threonine protein phosphatase. In an alternative example, SEQ ID NO:4 is 45% identical to human protein tyrosine phosphatase (GenBank ID g452194) as determined by the Basic Local Alignment Search Tool (BLAST). (See Table 2.) The BLAST probability score is 2.6e-169, which indicates the probability of obtaining the observed polypeptide sequence alignment by chance. SEQ ID NO:4 also contains a FERM domain (Band 4.1 family) as determined by searching for statistically significant matches in the hidden Markov model (HMM)-based PFAM database of conserved protein family domains. (See Table 3.) Data from BLIMPS and PROFILESCAN analyses provide further corroborative evidence that SEQ ID NO:4 contains a Band 4.1 family domain which is found in protein tyrosine phosphatases (note that the "Band 4.1 family domain signatures" is a conserved N-terminal domain of about 150 amino-acid residues known to exist in protein tyrosine phosphatases and could act at junctions between the plasma membrane and the cytoskeleton (Rees,D.J.G. et al., (1990) Nature 347:685-689, Funayama,N. et al., (1991) J. Cell Biol. 115:1039-1048, and Q. Yang and N.K. Tonks (1991) Proc. Natl. Acad. Sci. U.S.A. 88:5949-5953). In another alternative example, SEQ ID NO:7 is 57% identical to *Drosophila melanogaster* MAP kinase phosphatase (GenBank ID g6714641) as determined by the Basic Local Alignment Search Tool (BLAST). (See Table 2.) The BLAST probability score is 7.3e-101, which indicates the probability of obtaining the observed polypeptide sequence alignment by chance. SEQ ID NO:7 also contains a dual specificity phosphatase catalytic domain as determined by searching for statistically significant matches in the hidden Markov model (HMM)-based PFAM database of conserved protein family domains. (See Table 3.) Data from BLIMPS analysis provides further corroborative evidence that SEQ ID NO:7 is a dual-specificity phosphatase. In another alternative example, SEQ ID NO:9 is 46% identical to bovine protein phosphatase 2C beta (GenBank ID g3063745) as determined by the Basic Local

Alignment Search Tool (BLAST). (See Table 2.) The BLAST probability score is 3.5e-77, which indicates the probability of obtaining the observed polypeptide sequence alignment by chance. SEQ ID NO:9 also contains a protein phosphatase 2C proteins domain as determined by searching for statistically significant matches in the hidden Markov model (HMM)-based PFAM database of conserved protein family domains. (See Table 3.) Data from BLIMPS and MOTIFS analyses provide further corroborative evidence that SEQ ID NO:9 is a protein phosphatase 2C. In another alternative example, SEQ ID NO:11 has 97% local identity to human striatum-enriched phosphatase (GenBank ID g957217) as determined by the Basic Local Alignment Search Tool (BLAST). (See Table 2.) The BLAST probability score is 2.8e-292, which indicates the probability of obtaining the observed polypeptide sequence alignment by chance. SEQ ID NO:11 also contains a tyrosine phosphatase active site domain as determined by searching for statistically significant matches in the hidden Markov model (HMM)-based PFAM database of conserved protein family domains. (See Table 3.) Data from BLIMPS, MOTIFS, and PROFILESCAN analyses provide further corroborative evidence that SEQ ID NO:11 is a tyrosine specific phosphatase. In another alternative example, SEQ ID NO:12 is 1511 amino acids in length and is 99% identical over 1441 residues to human synaptojanin 2B (GenBank ID g4104822) as determined by the Basic Local Alignment Search Tool (BLAST). (See Table 2.) The BLAST probability score is 0.0, which indicates the probability of obtaining the observed polypeptide sequence alignment by chance. SEQ ID NO:12 also contains an inositol polyphosphate phosphatase family catalytic domain as determined by searching for statistically significant matches in the hidden Markov model (HMM)-based PFAM database of conserved protein family domains. (See Table 3.) Data from BLIMPS analyses provide further corroborative evidence that SEQ ID NO:12 is a synaptojanin (note that "synaptojanin" is a specific subfamily of the primary family of "protein phosphatases"). SEQ ID NO:1, SEQ ID NO:3, SEQ ID NO:5-6, SEQ ID NO:8 and SEQ ID NO:10 were analyzed and annotated in a similar manner. The algorithms and parameters for the analysis of SEQ ID NO:1-12 are described in Table 7.

As shown in Table 4, the full length polynucleotide sequences of the present invention were assembled using cDNA sequences or coding (exon) sequences derived from genomic DNA, or any combination of these two types of sequences. Columns 1 and 2 list the polynucleotide sequence identification number (Polynucleotide SEQ ID NO:) and the corresponding Incyte polynucleotide consensus sequence number (Incyte Polynucleotide ID) for each polynucleotide of the invention. Column 3 shows the length of each polynucleotide sequence in basepairs. Column 4 lists fragments of the polynucleotide sequences which are useful, for example, in hybridization or amplification technologies that identify SEQ ID NO:13-24 or that distinguish between SEQ ID NO:13-24 and related polynucleotide sequences. Column 5 shows identification numbers corresponding to cDNA sequences, coding sequences (exons) predicted from genomic DNA, and/or sequence assemblages comprised of both cDNA and genomic DNA. These sequences were used to assemble the full length polynucleotide sequences of the invention. Columns 6 and 7 of Table 4 show the nucleotide start (5′) and stop (3′) positions of the cDNA and/or genomic sequences in column 5 relative to their respective full length sequences.

The identification numbers in Column 5 of Table 4 may refer specifically, for example, to Incyte cDNAs along with their corresponding cDNA libraries. For example, 2013147H1 is the identification number of an Incyte cDNA sequence, and TESTNOT03 is the cDNA library from which it is derived. Incyte cDNAs for which cDNA libraries are not indicated were derived from pooled cDNA libraries (e.g., 71163473V1). Alternatively, the identification numbers in column 5 may refer to GenBank cDNAs or ESTs (e.g., g3163696) which contributed to the assembly of the full length polynucleotide sequences. In addition, the identification numbers in column 5 may identify sequences derived from the ENSEMBL (The Sanger Centre, Cambridge, UK) database (*i.e.*, those sequences including the designation "ENST"). Alternatively, the identification numbers in column 5 may be derived from the NCBI RefSeq Nucleotide Sequence Records Database (*i.e.*, those sequences including the designation "NM" or "NT") or the NCBI RefSeq Protein Sequence Records (*i.e.*, those sequences including the designation "NP"). Alternatively, the identification numbers in column 5 may refer to assemblages of both cDNA and Genscan-predicted exons brought together by an "exon stitching" algorithm. Alternatively, the identification numbers in column 5 may refer to assemblages of exons brought together by an "exon-stretching" algorithm. In instances where a RefSeq sequence was used as a protein homolog for the "exon-stretching" algorithm, a RefSeq identifier (denoted by "NM," "NP," or "NT") may be used in place of the GenBank identifier (*i.e.*, g*BBBBB*).

Alternatively, a prefix identifies component sequences that were hand-edited, predicted from genomic DNA sequences, or derived from a combination of sequence analysis methods. The following Table lists examples of component sequence prefixes and corresponding sequence analysis methods associated with the prefixes (see Example IV and Example V).

Prefix	**Type of analysis and/or examples of programs**
GNN, GFG, ENST	Exon prediction from genomic sequences using, for example, GENSCAN (Stanford University, CA, USA) or FGENES (Computer Genomics Group, The Sanger Centre, Cambridge, UK).
GBI	Hand-edited analysis of genomic sequences.
FL	Stitched or stretched genomic sequences (see Example V).
INCY	Full length transcript and exon prediction from mapping of EST sequences to the genome. Genomic location and EST composition data are combined to predict the exons and resulting transcript.

In some cases, Incyte cDNA coverage redundant with the sequence coverage shown in column 5 was obtained to confirm the final consensus polynucleotide sequence, but the relevant Incyte cDNA identification numbers are not shown.

Table 5 shows the representative cDNA libraries for those full length polynucleotide sequences which were assembled using Incyte cDNA sequences. The representative cDNA library is the Incyte cDNA library which is most frequently represented by the Incyte cDNA sequences which were used to assemble and confirm the above polynucleotide sequences. The tissues and vectors which were used to construct the cDNA libraries shown in Table 5 are described in Table 6.

The invention also encompasses PP variants. A preferred PP variant is one which has at least about 80%, or alternatively at least about 90%, or even at least about 95% amino acid sequence identity to the PP amino acid sequence, and which contains at least one functional or structural characteristic of PP.

The invention also encompasses polynucleotides which encode PP. In a particular embodiment, the invention encompasses a polynucleotide sequence comprising a sequence selected from the group consisting of SEQ ID NO:13-24, which encodes PP. The polynucleotide sequences of SEQ ID NO:13-24, as presented in the Sequence Listing, embrace the equivalent RNA sequences, wherein occurrences of the nitrogenous base thymine are replaced with uracil, and the sugar backbone is composed of ribose instead of deoxyribose.

The invention also encompasses a variant of a polynucleotide sequence encoding PP. In particular, such a variant polynucleotide sequence will have at least about 70%, or alternatively at least about 85%, or even at least about 95% polynucleotide sequence identity to the polynucleotide sequence encoding PP. A particular aspect of the invention encompasses a variant of a polynucleotide sequence comprising a sequence selected from the group consisting of SEQ ID NO:13-24 which has at least about 70%, or alternatively at least about 85%, or even at least about 95% polynucleotide sequence identity to a nucleic acid sequence selected from the group consisting of SEQ ID NO:13-24. Any one of the polynucleotide variants described above can encode an amino acid sequence which contains at least one functional or structural characteristic of PP.

It will be appreciated by those skilled in the art that as a result of the degeneracy of the genetic code, a multitude of polynucleotide sequences encoding PP, some bearing minimal similarity to the polynucleotide sequences of any known and naturally occurring gene, may be produced. Thus, the invention contemplates each and every possible variation of polynucleotide sequence that could be made by selecting combinations based on possible codon choices. These combinations are made in accordance with the standard triplet genetic code as applied to the polynucleotide sequence of naturally occurring PP, and all such variations are to be considered as being specifically disclosed.

Although nucleotide sequences which encode PP and its variants are generally capable of hybridizing to the nucleotide sequence of the naturally occurring PP under appropriately selected conditions of stringency, it may be advantageous to produce nucleotide sequences encoding PP or its derivatives possessing a substantially different codon usage, e.g., inclusion of non-naturally occurring codons. Codons may be selected to increase the rate at which expression of the peptide occurs in a particular prokaryotic or eukaryotic host in accordance with the frequency with which particular codons are utilized by the host. Other reasons for substantially altering the nucleotide sequence encoding PP and its derivatives without altering the encoded amino acid sequences include the production of RNA transcripts having more desirable properties, such as a greater half-life, than transcripts produced from the naturally occurring sequence.

The invention also encompasses production of DNA sequences which encode PP and PP derivatives, or fragments thereof, entirely by synthetic chemistry. After production, the synthetic sequence may be inserted into any of the many available expression vectors and cell systems using reagents well known in the art. Moreover, synthetic chemistry may be used to introduce mutations into a sequence encoding PP or any fragment thereof.

Also encompassed by the invention are polynucleotide sequences that are capable of hybridizing to the claimed polynucleotide sequences, and, in particular, to those shown in SEQ ID NO:13-24 and fragments thereof under various conditions of stringency. (See, e.g., Wahl, G.M. and S.L. Berger (1987) Methods Enzymol. 152:399-407; Kimmel, A.R. (1987) Methods Enzymol. 152:507-511.) Hybridization conditions, including annealing and wash conditions, are described in "Definitions."

Methods for DNA sequencing are well known in the art and may be used to practice any of the embodiments of the invention. The methods may employ such enzymes as the Klenow fragment of DNA polymerase I, SEQUENASE (US Biochemical, Cleveland OH), Taq polymerase (Applied Biosystems), thermostable T7 polymerase (Amersham Pharmacia Biotech, Piscataway NJ), or combinations of polymerases and proofreading exonucleases such as those found in the ELONGASE amplification system (Life Technologies, Gaithersburg MD). Preferably, sequence preparation is automated with machines such as the MICROLAB 2200 liquid transfer system (Hamilton, Reno NV), PTC200 thermal cycler (MJ Research, Watertown MA) and ABI CATALYST 800 thermal cycler (Applied Biosystems). Sequencing is then carried out using either the ABI 373 or 377 DNA sequencing system (Applied Biosystems), the MEGABACE 1000 DNA sequencing system (Molecular Dynamics, Sunnyvale CA), or other systems known in the art. The resulting sequences are analyzed using a variety of algorithms which are well known in the art. (See, e.g., Ausubel, F.M. (1997) Short Protocols in Molecular Biology, John Wiley & Sons, New York NY, unit 7.7; Meyers, R.A. (1995) Molecular Biology and Biotechnology, Wiley VCH, New York NY, pp. 856-853.)

The nucleic acid sequences encoding PP may be extended utilizing a partial nucleotide sequence and employing various PCR-based methods known in the art to detect upstream sequences, such as promoters and regulatory elements. For example, one method which may be employed, restriction-site PCR, uses universal and nested primers to amplify unknown sequence from genomic DNA within a cloning vector. (See, e.g., Sarkar, G. (1993) PCR Methods Applic. 2:318-322.) Another method, inverse PCR, uses primers that extend in divergent directions to amplify unknown sequence from a circularized template. The template is derived from restriction fragments comprising a known genomic locus and surrounding sequences. (See, e.g., Triglia, T. et al. (1988) Nucleic Acids Res. 16:8186.) A third method, capture PCR, involves PCR amplification of DNA fragments adjacent to known sequences in human and yeast artificial chromosome DNA. (See, e.g., Lagerstrom, M. et al. (1991) PCR Methods Applic. 1:111-119.) In this method, multiple restriction enzyme digestions and ligations may be used to insert an engineered double-stranded sequence into a region of unknown sequence before performing PCR. Other methods which may be used to retrieve unknown sequences are known in the art. (See, e.g., Parker, J.D. et al. (1991) Nucleic Acids Res. 19:3055-3060). Additionally, one may use PCR, nested primers, and PROMOTERFINDER libraries (Clontech, Palo Alto CA) to walk genomic DNA. This procedure avoids the need to screen libraries and is useful in finding intron/exon junctions. For all PCR-based methods, primers may be designed using commercially available software, such as OLIGO 4.06 primer analysis software (National Biosciences, Plymouth MN) or another appropriate program, to be about 22 to 30 nucleotides in length, to have a GC content of about 50% or more, and to anneal to the template at temperatures of about 68°C to 72°C.

When screening for full length cDNAs, it is preferable to use libraries that have been size-selected to include larger cDNAs. In addition, random-primed libraries, which often include sequences containing the 5' regions of genes, are preferable for situations in which an oligo d(T) library does not yield a full-length cDNA. Genomic libraries may be useful for extension of sequence into 5' non-transcribed regulatory regions.

Capillary electrophoresis systems which are commercially available may be used to analyze the size or confirm the nucleotide sequence of sequencing or PCR products. In particular, capillary sequencing may employ flowable polymers for electrophoretic separation, four different nucleotide-specific, laser-stimulated fluorescent dyes, and a charge coupled device camera for detection of the emitted wavelengths. Output/light intensity may be converted to electrical signal using appropriate software (e.g., GENOTYPER and SEQUENCE NAVIGATOR, Applied Biosystems), and the entire process from loading of samples to computer analysis and electronic data display may be computer controlled. Capillary electrophoresis is especially preferable for sequencing small DNA fragments which may be present in limited amounts in a particular sample.

In another embodiment of the invention, polynucleotide sequences or fragments thereof which encode PP may be cloned in recombinant DNA molecules that direct expression of PP, or fragments or functional equivalents thereof, in appropriate host cells. Due to the inherent degeneracy of the genetic code, other DNA sequences which encode substantially the same or a functionally equivalent amino acid sequence may be produced and used to express PP.

The nucleotide sequences of the present invention can be engineered using methods generally known in the art in order to alter PP-encoding sequences for a variety of purposes including, but not limited to, modification of the cloning, processing, and/or expression of the gene product. DNA shuffling by random fragmentation and PCR reassembly of gene fragments and synthetic oligonucleotides may be used to engineer the nucleotide sequences. For example, oligonucleotide-mediated site-directed mutagenesis may be used to introduce mutations that create new restriction sites, alter glycosylation patterns, change codon preference, produce splice variants, and so forth.

The nucleotides of the present invention may be subjected to DNA shuffling techniques such as MOLECULARBREEDING (Maxygen Inc., Santa Clara CA; described in U.S. Patent Number 5,837,458; Chang, C.-C. et al. (1999) Nat. Biotechnol. 17:793-797; Christians, F.C. et al. (1999) Nat. Biotechnol. 17:259-264; and Crameri, A. et al. (1996) Nat. Biotechnol. 14:315-319) to alter or improve the biological properties of PP, such as its biological or enzymatic activity or its ability to bind to other molecules or compounds. DNA shuffling is a process by which a library of gene variants is produced using PCR-mediated recombination of gene fragments. The library is then subjected to selection or screening procedures that identify those gene variants with the desired properties. These preferred variants may then be pooled and further subjected to recursive rounds of DNA shuffling and selection/screening. Thus, genetic diversity is created through "artificial" breeding and rapid molecular evolution. For example, fragments of a single gene containing random point mutations may be recombined, screened, and then reshuffled until the desired properties are optimized. Alternatively, fragments of a given gene may be recombined with fragments of homologous genes in the same gene family, either from the same or different species, thereby maximizing the genetic diversity of multiple naturally occurring genes in a directed and controllable manner.

In another embodiment, sequences encoding PP may be synthesized, in whole or in part, using chemical methods well known in the art. (See, e.g., Caruthers, M.H. et al. (1980) Nucleic Acids Symp. Ser. 7:215-223; and Horn, T. et al. (1980) Nucleic Acids Symp. Ser. 7:225-232.) Alternatively, PP itself or a fragment thereof may be synthesized using chemical methods. For example, peptide synthesis can be performed using various solution-phase or solid-phase techniques. (See, e.g., Creighton, T. (1984) Proteins, Structures and Molecular Properties, WH Freeman, New York NY, pp. 55-60; and Roberge, J.Y. et al. (1995) Science 269:202-204.) Automated synthesis may be achieved using the ABI 431A peptide synthesizer (Applied Biosystems). Additionally, the amino acid sequence of PP, or any part thereof, may be altered during direct synthesis and/or combined with sequences from other proteins, or any part thereof, to produce a variant polypeptide or a polypeptide having a sequence of a naturally occurring polypeptide.

The peptide may be substantially purified by preparative high performance liquid chromatography. (See, e.g., Chiez, R.M. and F.Z. Regnier (1990) Methods Enzymol. 182:392-421.) The composition of the synthetic peptides may be confirmed by amino acid analysis or by sequencing. (See, e.g., Creighton, supra, pp. 28-53.)

In order to express a biologically active PP, the nucleotide sequences encoding PP or derivatives thereof may be inserted into an appropriate expression vector, i.e., a vector which contains the necessary elements for transcriptional and translational control of the inserted coding sequence in a suitable host. These elements include regulatory sequences, such as enhancers, constitutive and inducible promoters, and 5' and 3' untranslated regions in the vector and in polynucleotide sequences encoding PP. Such elements may vary in their strength and specificity. Specific initiation signals may also be used to achieve more efficient translation of sequences encoding PP. Such signals include the ATG initiation codon and adjacent sequences, e.g. the Kozak sequence. In cases where sequences encoding PP and its initiation codon and upstream regulatory sequences are inserted into the appropriate expression vector, no additional transcriptional or translational control signals may be needed. However, in cases where only coding sequence, or a fragment thereof, is inserted, exogenous translational control signals including an in-frame ATG initiation codon should be provided by the vector. Exogenous translational elements and initiation codons may be of various origins, both natural and synthetic. The efficiency of expression may be enhanced by the inclusion of enhancers appropriate for the particular host cell system used. (See, e.g., Scharf, D. et al. (1994) Results Probl. Cell Differ. 20:125-162.)

Methods which are well known to those skilled in the art may be used to construct expression vectors containing sequences encoding PP and appropriate transcriptional and translational control elements. These methods include in vitro recombinant DNA techniques, synthetic techniques, and in vivo genetic recombination. (See, e.g., Sambrook, J. et al. (1989) Molecular Cloning, A Laboratory Manual, Cold Spring Harbor Press, Plainview NY, ch. 4, 8, and 16-17; Ausubel, F.M. et al. (1995) Current Protocols in Molecular Biology, John Wiley & Sons, New York NY, ch. 9, 13, and 16.)

A variety of expression vector/host systems may be utilized to contain and express sequences encoding PP. These include, but are not limited to, microorganisms such as bacteria transformed with recombinant bacteriophage, plasmid, or cosmid DNA expression vectors; yeast transformed with yeast expression vectors; insect cell systems infected with viral expression vectors (e.g., baculovirus); plant cell systems transformed with viral expression vectors (e.g., cauliflower mosaic virus, CaMV, or tobacco mosaic virus, TMV) or with bacterial expression vectors (e.g., Ti or pBR322 plasmids); or animal cell systems. (See, e.g., Sambrook, supra; Ausubel, supra; Van Heeke, G. and S.M. Schuster (1989) J. Biol. Chem. 264:5503-5509; Engelhard, E.K. et al. (1994) Proc. Natl. Acad. Sci. USA 91:3224-3227; Sandig, V. et al. (1996) Hum. Gene Ther. 7:1937-1945; Takamatsu, N. (1987) EMBO J. 6:307-311; The McGraw Hill Yearbook of Science and Technology (1992) McGraw Hill, New York NY, pp. 191-196; Logan, J. and T. Shenk (1984) Proc. Natl. Acad. Sci. USA 81:3655-3659; and Harrington, J.J. et al. (1997) Nat. Genet. 15:345-355.) Expression vectors derived from retroviruses, adenoviruses, or herpes or vaccinia viruses, or from various bacterial plasmids, may be used for delivery of nucleotide sequences to the targeted organ, tissue, or cell population. (See, e.g., Di Nicola, M. et al. (1998) Cancer Gen. Ther. 5(6):350-356; Yu, M. et al. (1993) Proc. Natl. Acad. Sci. USA 90(13):6340-6344; Buller, R.M. et al. (1985) Nature 317(6040):813-815; McGregor, D.P. et al. (1994) Mol. Immunol. 31(3):219-226; and Verma, I.M. and N. Somia (1997) Nature 389:239-242.) The invention is not limited by the host cell employed.

In bacterial systems, a number of cloning and expression vectors may be selected depending upon the use intended for polynucleotide sequences encoding PP. For example, routine cloning, subcloning, and propagation of polynucleotide sequences encoding PP can be achieved using a multifunctional E. coli vector such as PBLUESCRIPT (Stratagene, La Jolla CA) or PSPORT1 plasmid (Life Technologies). Ligation of sequences encoding PP into the vector's multiple cloning site disrupts the *lac*Z gene, allowing a colorimetric screening procedure for identification of transformed bacteria containing recombinant molecules. In addition, these vectors may be useful for in vitro transcription, dideoxy sequencing, single strand rescue with helper phage, and creation of nested deletions in the cloned sequence. (See, e.g., Van Heeke, G. and S.M. Schuster (1989) J. Biol. Chem. 264:5503-5509.) When large quantities of PP are needed, e.g. for the production of antibodies, vectors which direct high level expression of PP may be used. For example, vectors containing the strong, inducible SP6 or T7 bacteriophage promoter may be used.

Yeast expression systems may be used for production of PP. A number of vectors containing constitutive or inducible promoters, such as alpha factor, alcohol oxidase, and PGH promoters, may be used in the yeast Saccharomyces cerevisiae or Pichia pastoris. In addition, such vectors direct either the secretion or intracellular retention of expressed proteins and enable integration of foreign sequences into the host genome for stable propagation. (See, e.g., Ausubel, 1995, supra; Bitter, G.A. et al. (1987) Methods Enzymol. 153:516-544; and Scorer, C.A. et al. (1994) Bio/Technology 12:181-184.)

Plant systems may also be used for expression of PP. Transcription of sequences encoding PP may be driven by viral promoters, e.g., the 35S and 19S promoters of CaMV used alone or in combination with the omega leader sequence from TMV (Takamatsu, N. (1987) EMBO J.

6:307-311). Alternatively, plant promoters such as the small subunit of RUBISCO or heat shock promoters may be used. (See, e.g., Coruzzi, G. et al. (1984) EMBO J. 3:1671-1680; Broglie, R. et al. (1984) Science 224:838-843; and Winter, J. et al. (1991) Results Probl. Cell Differ. 17:85-105.) These constructs can be introduced into plant cells by direct DNA transformation or pathogen-mediated transfection. (See, e.g., The McGraw Hill Yearbook of Science and Technology (1992) McGraw Hill, New York NY, pp. 191-196.)

In mammalian cells, a number of viral-based expression systems may be utilized. In cases where an adenovirus is used as an expression vector, sequences encoding PP may be ligated into an adenovirus transcription/translation complex consisting of the late promoter and tripartite leader sequence. Insertion in a non-essential E1 or E3 region of the viral genome may be used to obtain infective virus which expresses PP in host cells. (See, e.g., Logan, J. and T. Shenk (1984) Proc. Natl. Acad. Sci. USA 81:3655-3659.) In addition, transcription enhancers, such as the Rous sarcoma virus (RSV) enhancer, may be used to increase expression in mammalian host cells. SV40 or EBV-based vectors may also be used for high-level protein expression.

Human artificial chromosomes (HACs) may also be employed to deliver larger fragments of DNA than can be contained in and expressed from a plasmid. HACs of about 6 kb to 10 Mb are constructed and delivered via conventional delivery methods (liposomes, polycationic amino polymers, or vesicles) for therapeutic purposes. (See, e.g., Harrington, J.J. et al. (1997) Nat. Genet. 15:345-355.)

For long term production of recombinant proteins in mammalian systems, stable expression of PP in cell lines is preferred. For example, sequences encoding PP can be transformed into cell lines using expression vectors which may contain viral origins of replication and/or endogenous expression elements and a selectable marker gene on the same or on a separate vector. Following the introduction of the vector, cells may be allowed to grow for about 1 to 2 days in enriched media before being switched to selective media. The purpose of the selectable marker is to confer resistance to a selective agent, and its presence allows growth and recovery of cells which successfully express the introduced sequences. Resistant clones of stably transformed cells may be propagated using tissue culture techniques appropriate to the cell type.

Any number of selection systems may be used to recover transformed cell lines. These include, but are not limited to, the herpes simplex virus thymidine kinase and adenine phosphoribosyltransferase genes, for use in tk^- and apr^- cells, respectively. (See, e.g., Wigler, M. et al. (1977) Cell 11:223-232; Lowy, I. et al. (1980) Cell 22:817-823.) Also, antimetabolite, antibiotic, or herbicide resistance can be used as the basis for selection. For example, *dhfr* confers resistance to methotrexate; *neo* confers resistance to the aminoglycosides neomycin and G-418; and *als* and *pat* confer resistance to chlorsulfuron and phosphinotricin acetyltransferase, respectively. (See, e.g., Wigler, M. et al. (1980) Proc. Natl. Acad. Sci. USA 77:3567-3570; Colbere-Garapin, F. et al. (1981) J. Mol. Biol. 150:1-14.) Additional selectable genes have been described, e.g., *trpB* and *hisD*, which alter cellular requirements for metabolites. (See, e.g., Hartman, S.C. and R.C. Mulligan (1988) Proc. Natl. Acad. Sci. USA 85:8047-8051.) Visible markers, e.g., anthocyanins, green fluorescent proteins (GFP; Clontech), ß glucuronidase and its substrate ß-glucuronide, or luciferase and its substrate luciferin may be used. These markers can be used not only to identify transformants, but also to quantify the amount of transient or stable protein expression attributable to a specific vector system. (See, e.g., Rhodes, C.A. (1995) Methods Mol. Biol. 55:121-131.)

Although the presence/absence of marker gene expression suggests that the gene of interest is also present, the presence and expression of the gene may need to be confirmed. For example, if the sequence encoding PP is inserted within a marker gene sequence, transformed cells containing sequences encoding PP can be identified by the absence of marker gene function. Alternatively, a marker gene can be placed in tandem with a sequence encoding PP under the control of a single promoter. Expression of the marker gene in response to induction or selection usually indicates expression of the tandem gene as well.

In general, host cells that contain the nucleic acid sequence encoding PP and that express PP may be identified by a variety of procedures known to those of skill in the art. These procedures include, but are not limited to, DNA-DNA or DNA-RNA hybridizations, PCR amplification, and protein bioassay or immunoassay techniques which include membrane, solution, or chip based technologies for the detection and/or quantification of nucleic acid or protein sequences.

Immunological methods for detecting and measuring the expression of PP using either specific polyclonal or monoclonal antibodies are known in the art. Examples of such techniques include enzyme-linked immunosorbent assays (ELISAs), radioimmunoassays (RIAs), and fluorescence activated cell sorting (FACS). A two-site, monoclonal-based immunoassay utilizing monoclonal antibodies reactive to two non-interfering epitopes on PP is preferred, but a competitive binding assay may be employed. These and other assays are well known in the art. (See, e.g., Hampton, R. et al. (1990) Serological Methods, a Laboratory Manual, APS Press, St. Paul MN, Sect. IV; Coligan, J.E. et al. (1997) Current Protocols in Immunology, Greene Pub. Associates and Wiley-Interscience, New York NY; and Pound, J.D. (1998) Immunochemical Protocols, Humana Press, Totowa NJ.)

A wide variety of labels and conjugation techniques are known by those skilled in the art and may be used in various nucleic acid and amino acid assays. Means for producing labeled hybridization or PCR probes for detecting sequences related to polynucleotides encoding PP include oligolabeling, nick translation, end-labeling, or PCR amplification using a labeled nucleotide. Alternatively, the sequences encoding PP, or any fragments thereof, may be cloned into a vector for the production of an mRNA probe. Such vectors are known in the art, are commercially available, and may be used to synthesize RNA probes in vitro by addition of an appropriate RNA polymerase such as T7, T3, or SP6 and labeled nucleotides. These procedures may be conducted using a variety of commercially available kits, such as those provided by Amersham Pharmacia Biotech, Promega (Madison WI), and US Biochemical. Suitable reporter molecules or labels which may be used for ease of detection include radionuclides, enzymes, fluorescent, chemiluminescent, or chromogenic agents, as well as substrates, cofactors, inhibitors, magnetic particles, and the like.

Host cells transformed with nucleotide sequences encoding PP may be cultured under conditions suitable for the expression and recovery of the protein from cell culture. The protein produced by a transformed cell may be secreted or retained intracellularly depending on the sequence and/or the vector used. As will be understood by those of skill in the art, expression vectors containing polynucleotides which encode PP may be designed to contain signal sequences which direct secretion of PP through a prokaryotic or eukaryotic cell membrane.

In addition, a host cell strain may be chosen for its ability to modulate expression of the inserted sequences or to process the expressed protein in the desired fashion. Such modifications of the polypeptide include, but are not limited to, acetylation, carboxylation, glycosylation, phosphorylation, lipidation, and acylation. Post-translational processing which cleaves a "prepro" or "pro" form of the protein may also be used to specify protein targeting, folding, and/or activity. Different host cells which have specific cellular machinery and characteristic mechanisms for post-translational activities (e.g., CHO, HeLa, MDCK, HEK293, and WI38) are available from the American Type Culture Collection (ATCC, Manassas VA) and may be chosen to ensure the correct modification and processing of the foreign protein.

In another embodiment of the invention, natural, modified, or recombinant nucleic acid sequences encoding PP may be ligated to a heterologous sequence resulting in translation of a fusion protein in any of the aforementioned host systems. For example, a chimeric PP protein containing a heterologous moiety that can be recognized by a commercially available antibody may facilitate the screening of peptide libraries for inhibitors of PP activity. Heterologous protein and peptide moieties may also facilitate purification of fusion proteins using commercially available affinity matrices. Such moieties include, but are not limited to, glutathione S-transferase (GST), maltose binding protein (MBP), thioredoxin (Trx), calmodulin binding peptide (CBP), 6-His, FLAG, *c-myc*, and hemagglutinin (HA). GST, MBP, Trx, CBP, and 6-His enable purification of their cognate fusion proteins on immobilized glutathione, maltose, phenylarsine oxide, calmodulin, and metal-chelate resins, respectively. FLAG, *c-myc*, and hemagglutinin (HA) enable immunoaffinity purification of fusion proteins using commercially available monoclonal and polyclonal antibodies that specifically recognize these epitope tags. A fusion protein may also be engineered to contain a proteolytic cleavage site located between the PP encoding sequence and the heterologous protein sequence, so that PP may be cleaved away from the heterologous moiety following purification. Methods for fusion protein expression and purification are discussed in Ausubel (1995, supra, ch. 10). A variety of commercially available kits may also be used to facilitate expression and purification of fusion proteins.

In a further embodiment of the invention, synthesis of radiolabeled PP may be achieved in vitro using the TNT rabbit reticulocyte lysate or wheat germ extract system (Promega). These systems couple transcription and translation of protein-coding sequences operably associated with the T7, T3, or SP6 promoters. Translation takes place in the presence of a radiolabeled amino acid precursor, for example, ^{35}S-methionine.

PP of the present invention or fragments thereof may be used to screen for compounds that specifically bind to PP. At least one and up to a plurality of test compounds may be screened for specific binding to PP. Examples of test compounds include antibodies, oligonucleotides, proteins (e.g., receptors), or small molecules.

In one embodiment, the compound thus identified is closely related to the natural ligand of PP, e.g., a ligand or fragment thereof, a natural substrate, a structural or functional mimetic, or a natural binding partner. (See, e.g., Coligan, J.E. et al. (1991) Current Protocols in Immunology 1(2): Chapter 5.) Similarly, the compound can be closely related to the natural receptor to which PP binds, or to at least a fragment of the receptor, e.g., the ligand binding site. In either case, the compound can be rationally designed using known techniques. In one embodiment, screening for these compounds involves producing appropriate cells which express PP, either as a secreted protein or on the cell membrane. Preferred cells include cells from mammals, yeast, Drosophila, or E. coli. Cells expressing PP or cell membrane fractions which contain PP are then contacted with a test compound and binding, stimulation, or inhibition of activity of either PP or the compound is analyzed.

An assay may simply test binding of a test compound to the polypeptide, wherein binding is detected by a fluorophore, radioisotope, enzyme conjugate, or other detectable label. For example, the assay may comprise the steps of combining at least one test compound with PP, either in solution or affixed to a solid support, and detecting the binding of PP to the compound. Alternatively, the assay may detect or measure binding of a test compound in the presence of a labeled competitor. Additionally, the assay may be carried out using cell-free preparations, chemical libraries, or natural product mixtures, and the test compound(s) may be free in solution or affixed to a solid support.

PP of the present invention or fragments thereof may be used to screen for compounds that modulate the activity of PP. Such compounds may include agonists, antagonists, or partial or inverse agonists. In one embodiment, an assay is performed under conditions permissive for PP activity, wherein PP is combined with at least one test compound, and the activity of PP in the presence of a test compound is compared with the activity of PP in the absence of the test compound. A change in the activity of PP in the presence of the test compound is indicative of a compound that modulates the activity of PP. Alternatively, a test compound is combined with an in vitro or cell-free system comprising PP under conditions suitable for PP activity, and the assay is performed. In either of these assays, a test compound which modulates the activity of PP may do so indirectly and need not come in direct contact with the test compound. At least one and up to a plurality of test compounds may be screened.

In another embodiment, polynucleotides encoding PP or their mammalian homologs may be "knocked out" in an animal model system using homologous recombination in embryonic stem (ES) cells. Such techniques are well known in the art and are useful for the generation of animal models of human disease. (See, e.g., U.S. Patent Number 5,175,383 and U.S. Patent Number 5,767,337.) For example, mouse ES cells, such as the mouse 129/SvJ cell line, are derived from the early mouse embryo and grown in culture. The ES cells are transformed with a vector containing the gene of interest disrupted by a marker gene, e.g., the neomycin phosphotransferase gene (neo; Capecchi, M.R. (1989) Science 244:1288-1292). The vector integrates into the corresponding region of the host genome by homologous recombination. Alternatively, homologous recombination takes place using the Cre-loxP system to knockout a gene of interest in a tissue- or developmental stage-specific manner (Marth, J.D. (1996) Clin. Invest. 97:1999-2002; Wagner, K.U. et al. (1997) Nucleic Acids Res. 25:4323-4330). Transformed ES cells are identified and microinjected into mouse cell blastocysts such as those from the C57BL/6 mouse strain. The blastocysts are surgically transferred to pseudopregnant dams, and the resulting chimeric progeny are genotyped and bred to produce heterozygous or homozygous strains. Transgenic animals thus generated may be tested with potential therapeutic or toxic agents.

Polynucleotides encoding PP may also be manipulated in vitro in ES cells derived from human blastocysts. Human ES cells have the potential to differentiate into at least eight separate cell lineages including endoderm, mesoderm, and ectodermal cell types. These cell lineages differentiate into, for example, neural cells, hematopoietic lineages, and cardiomyocytes (Thomson, J.A. et al. (1998) Science 282:1145-1147).

Polynucleotides encoding PP can also be used to create "knockin" humanized animals (pigs) or transgenic animals (mice or rats) to model human disease. With knockin technology, a region of a polynucleotide encoding PP is injected into animal ES cells, and the injected sequence integrates into the animal cell genome. Transformed cells are injected into blastulae, and the blastulae are implanted as described above. Transgenic progeny or inbred lines are studied and treated with potential pharmaceutical agents to obtain information on treatment of a human disease. Alternatively, a mammal inbred to overexpress PP, e.g., by secreting PP in its milk, may also serve as a convenient source of that protein (Janne, J. et al. (1998) Biotechnol. Annu. Rev. 4:55-74).

THERAPEUTICS

Chemical and structural similarity, e.g., in the context of sequences and motifs, exists between regions of PP and protein phosphatases. In addition, the expression of PP is closely associated with bone, ovary, brain, prostate, abdominal fat, nervous, gastrointestinal and diseased tissues. Therefore, PP appears to play a role in immune system disorders, neurological disorders, developmental disorders, and cell proliferative disorders, including cancer. In the treatment of disorders associated with increased PP expression or activity, it is desirable to decrease the expression or activity of PP. In the treatment of disorders associated with decreased PP expression or activity, it is desirable to increase the expression or activity of PP.

Therefore, in one embodiment, PP or a fragment or derivative thereof may be administered to a subject to treat or prevent a disorder associated with decreased expression or activity of PP. Examples of such disorders include, but are not limited to, an immune system disorder, such as acquired immunodeficiency syndrome (AIDS), X-linked agammaglobinemia of Bruton, common variable immunodeficiency (CVI), DiGeorge's syndrome (thymic hypoplasia), thymic dysplasia, isolated IgA deficiency, severe combined immunodeficiency disease (SCID), immunodeficiency with thrombocytopenia and eczema (Wiskott-Aldrich syndrome), Chediak-Higashi syndrome, chronic granulomatous diseases, hereditary angioneurotic edema, immunodeficiency associated with Cushing's disease, Addison's disease, adult respiratory distress syndrome, allergies, ankylosing spondylitis, amyloidosis, anemia, asthma, atherosclerosis, autoimmune hemolytic anemia, autoimmune thyroiditis, autoimmune polyendocrinopathy-candidiasis-ectodermal dystrophy (APECED), bronchitis, cholecystitis, contact dermatitis, Crohn's disease, atopic dermatitis, dermatomyositis, diabetes mellitus, emphysema, episodic lymphopenia with lymphocytotoxins, erythroblastosis fetalis, erythema nodosum, atrophic gastritis, glomerulonephritis, Goodpasture's syndrome, gout, Graves' disease, Hashimoto's thyroiditis, hypereosinophilia, irritable bowel syndrome, multiple sclerosis, myasthenia gravis, myocardial or pericardial inflammation, osteoarthritis, osteoporosis, pancreatitis, polymyositis, psoriasis, Reiter's syndrome, rheumatoid arthritis, scleroderma, Sjögren's syndrome, systemic anaphylaxis, systemic lupus erythematosus, systemic sclerosis, thrombocytopenic purpura, ulcerative colitis, uveitis, Werner syndrome, complications of cancer, hemodialysis, and extracorporeal circulation, viral, bacterial, fungal, parasitic, protozoal, and helminthic infections, and trauma; a neurological disorder, such as epilepsy, ischemic cerebrovascular disease, stroke, cerebral neoplasms, Alzheimer's disease, Pick's disease, Huntington's disease, dementia, Parkinson's disease and other extrapyramidal disorders, amyotrophic lateral sclerosis and other motor neuron disorders, progressive neural muscular atrophy, retinitis pigmentosa, hereditary ataxias, multiple sclerosis and other demyelinating diseases, bacterial and viral meningitis, brain abscess, subdural empyema, epidural abscess, suppurative intracranial thrombophlebitis, myelitis and radiculitis, viral central nervous system disease, prion diseases including kuru, Creutzfeldt-Jakob disease, and Gerstmann-Straussler-Scheinker syndrome, fatal familial insomnia, nutritional and metabolic diseases of the nervous system, neurofibromatosis, tuberous sclerosis, cerebelloretinal hemangioblastomatosis, encephalotrigeminal syndrome, mental retardation and other developmental disorders of the central nervous system including Down syndrome, cerebral palsy, neuroskeletal disorders, autonomic nervous system disorders, cranial nerve disorders, spinal cord diseases, muscular dystrophy and other neuromuscular disorders, peripheral nervous system disorders, dermatomyositis and polymyositis, inherited, metabolic, endocrine, and toxic myopathies, myasthenia gravis, periodic paralysis, mental disorders including mood, anxiety, and schizophrenic disorders, seasonal affective disorder (SAD), akathesia, amnesia, catatonia, diabetic neuropathy, tardive dyskinesia, dystonias, paranoid psychoses, postherpetic neuralgia, Tourette's disorder, progressive supranuclear palsy, corticobasal degeneration, and familial frontotemporal dementia; a developmental disorder, such as renal tubular acidosis, anemia, Cushing's syndrome, achondroplastic dwarfism, Duchenne and Becker muscular dystrophy, epilepsy, gonadal dysgenesis, WAGR syndrome (Wilms' tumor, aniridia, genitourinary abnormalities, and mental retardation), Smith-Magenis syndrome, myelodysplastic syndrome, hereditary mucoepithelial dysplasia, hereditary keratodermas, hereditary neuropathies such as Charcot-Marie-Tooth disease and neurofibromatosis, hypothyroidism, hydrocephalus, seizure disorders such as Syndenham's chorea and cerebral palsy, spina bifida, anencephaly, craniorachischisis, congenital glaucoma, cataract, and sensorineural hearing loss; and a cell proliferative disorder, such as actinic keratosis, arteriosclerosis, atherosclerosis, bursitis, cirrhosis, hepatitis, mixed connective tissue disease (MCTD), myelofibrosis, paroxysmal nocturnal hemoglobinuria, polycythemia vera, psoriasis, primary thrombocythemia, and cancers including adenocarcinoma, leukemia, lymphoma, melanoma, myeloma, sarcoma, teratocarcinoma, and, in particular, cancers of the adrenal gland, bladder, bone, bone marrow, brain, breast, cervix, gall bladder, ganglia, gastrointestinal tract, heart, kidney, liver, lung, muscle, ovary, pancreas, parathyroid, penis, prostate, salivary glands, skin, spleen, testis, thymus, thyroid, and uterus.

In another embodiment, a vector capable of expressing PP or a fragment or derivative thereof may be administered to a subject to treat or prevent a disorder associated with decreased expression or activity of PP including, but not limited to, those described above.

In a further embodiment, a composition comprising a substantially purified PP in conjunction with a suitable pharmaceutical carrier may be administered to a subject to treat or prevent a disorder associated with decreased expression or activity of PP including, but not limited to, those provided above.

In still another embodiment, an agonist which modulates the activity of PP may be administered to a subject to treat or prevent a disorder associated with decreased expression or activity of PP including, but not limited to, those listed above.

In a further embodiment, an antagonist of PP may be administered to a subject to treat or prevent a disorder associated with increased expression or activity of PP. Examples of such disorders include, but are not limited to, those immune system disorders, neurological disorders, developmental disorders, and cell proliferative disorders, including cancer described above. In one aspect, an antibody which specifically binds PP may be used directly as an antagonist or indirectly as a targeting or delivery mechanism for bringing a pharmaceutical agent to cells or tissues which express PP.

In an additional embodiment, a vector expressing the complement of the polynucleotide encoding PP may be administered to a subject to treat or prevent a disorder associated with increased expression or activity of PP including, but not limited to, those described above.

In other embodiments, any of the proteins, antagonists, antibodies, agonists, complementary sequences, or vectors of the invention may be administered in combination with other appropriate therapeutic agents. Selection of the appropriate agents for use in combination therapy may be made by one of ordinary skill in the art, according to conventional pharmaceutical principles. The combination of therapeutic agents may act synergistically to effect the treatment or prevention of the various disorders described above. Using this approach, one may be able to achieve therapeutic efficacy with lower dosages of each agent, thus reducing the potential for adverse side effects.

An antagonist of PP may be produced using methods which are generally known in the art. In particular, purified PP may be used to produce antibodies or to screen libraries of pharmaceutical agents to identify those which specifically bind PP. Antibodies to PP may also be generated using methods that are well known in the art. Such antibodies may include, but are not limited to, polyclonal, monoclonal, chimeric, and single chain antibodies, Fab fragments, and fragments produced by a Fab expression library. Neutralizing antibodies (i.e., those which inhibit dimer formation) are generally preferred for therapeutic use.

For the production of antibodies, various hosts including goats, rabbits, rats, mice, humans, and others may be immunized by injection with PP or with any fragment or oligopeptide thereof which has immunogenic properties. Depending on the host species, various adjuvants may be used to increase immunological response. Such adjuvants include, but are not limited to, Freund's, mineral gels such as aluminum hydroxide, and surface active substances such as lysolecithin, pluronic polyols, polyanions, peptides, oil emulsions, KLH, and dinitrophenol. Among adjuvants used in humans, BCG (bacilli Calmette-Guerin) and Corynebacterium parvum are especially preferable.

It is preferred that the oligopeptides, peptides, or fragments used to induce antibodies to PP have an amino acid sequence consisting of at least about 5 amino acids, and generally will consist of at least about 10 amino acids. It is also preferable that these oligopeptides, peptides, or fragments are identical to a portion of the amino acid sequence of the natural protein. Short stretches of PP amino acids may be fused with those of another protein, such as KLH, and antibodies to the chimeric molecule may be produced.

Monoclonal antibodies to PP may be prepared using any technique which provides for the production of antibody molecules by continuous cell lines in culture. These include, but are not limited to, the hybridoma technique, the human B-cell hybridoma technique, and the EBV-hybridoma technique. (See, e.g., Kohler, G. et al. (1975) Nature 256:495-497; Kozbor, D. et al. (1985) J. Immunol. Methods 81:31-42; Cote, R.J. et al. (1983) Proc. Natl. Acad. Sci. USA 80:2026-2030; and Cole, S.P. et al. (1984) Mol. Cell Biol. 62:109-120.)

In addition, techniques developed for the production of "chimeric antibodies," such as the splicing of mouse antibody genes to human antibody genes to obtain a molecule with appropriate antigen specificity and biological activity, can be used. (See, e.g., Morrison, S.L. et al. (1984) Proc. Natl. Acad. Sci. USA 81:6851-6855; Neuberger, M.S. et al. (1984) Nature 312:604-608; and Takeda, S. et al. (1985) Nature 314:452-454.) Alternatively, techniques described for the production of single chain antibodies may be adapted, using methods known in the art, to produce PP-specific single chain antibodies. Antibodies with related specificity, but of distinct idiotypic composition, may be generated by chain shuffling from random combinatorial immunoglobulin libraries. (See, e.g., Burton, D.R. (1991) Proc. Natl. Acad. Sci. USA 88:10134-10137.)

Antibodies may also be produced by inducing in vivo production in the lymphocyte population or by screening immunoglobulin libraries or panels of highly specific binding reagents as disclosed in the literature. (See, e.g., Orlandi, R. et al. (1989) Proc. Natl. Acad. Sci. USA 86:3833-3837; Winter, G. et al. (1991) Nature 349:293-299.)

Antibody fragments which contain specific binding sites for PP may also be generated. For example, such fragments include, but are not limited to, $F(ab')_2$ fragments produced by pepsin digestion of the antibody molecule and Fab fragments generated by reducing the disulfide bridges of the F(ab')2 fragments. Alternatively, Fab expression libraries may be constructed to allow rapid and easy identification of monoclonal Fab fragments with the desired specificity. (See, e.g., Huse, W.D. et al. (1989) Science 246:1275-1281.)

Various immunoassays may be used for screening to identify antibodies having the desired specificity. Numerous protocols for competitive binding or immunoradiometric assays using either polyclonal or monoclonal antibodies with established specificities are well known in the art. Such immunoassays typically involve the measurement of complex formation between PP and its specific antibody. A two-site, monoclonal-based immunoassay utilizing monoclonal antibodies reactive to two non-interfering PP epitopes is generally used, but a competitive binding assay may also be employed (Pound, supra).

Various methods such as Scatchard analysis in conjunction with radioimmunoassay techniques may be used to assess the affinity of antibodies for PP. Affinity is expressed as an association constant, K_a, which is defined as the molar concentration of PP-antibody complex divided by the molar concentrations of free antigen and free antibody under equilibrium conditions. The K_a determined for a preparation of polyclonal antibodies, which are heterogeneous in their affinities for multiple PP epitopes, represents the average affinity, or avidity, of the antibodies for PP. The K_a determined for a preparation of monoclonal antibodies, which are monospecific for a particular PP epitope, represents a true measure of affinity. High-affinity antibody preparations with K_a ranging from about 10^9 to 10^{12} L/mole are preferred for use in immunoassays in which the PP-antibody complex must withstand rigorous manipulations. Low-affinity antibody preparations with K_a ranging from about 10^6 to 10^7 L/mole are preferred for use in immunopurification and similar procedures which ultimately require dissociation of PP, preferably in active form, from the antibody (Catty, D. (1988) Antibodies, Volume I: A Practical Approach, IRL Press, Washington DC; Liddell, J.E. and A. Cryer (1991) A Practical Guide to Monoclonal Antibodies, John Wiley & Sons, New York NY).

The titer and avidity of polyclonal antibody preparations may be further evaluated to determine the quality and suitability of such preparations for certain downstream applications. For example, a polyclonal antibody preparation containing at least 1-2 mg specific antibody/ml, preferably 5-10 mg specific antibody/ml, is generally employed in procedures requiring precipitation of PP-antibody complexes. Procedures for evaluating antibody specificity, titer, and avidity, and guidelines for antibody quality and usage in various applications, are generally available. (See, e.g., Catty, supra, and Coligan et al. supra.)

In another embodiment of the invention, the polynucleotides encoding PP, or any fragment or complement thereof, may be used for therapeutic purposes. In one aspect, modifications of gene expression can be achieved by designing complementary sequences or antisense molecules (DNA, RNA, PNA, or modified oligonucleotides) to the coding or regulatory regions of the gene encoding PP. Such technology is well known in the art, and antisense oligonucleotides or larger fragments can be designed from various locations along the coding or control regions of sequences encoding PP. (See, e.g., Agrawal, S., ed. (1996) Antisense Therapeutics, Humana Press Inc., Totawa NJ.)

In therapeutic use, any gene delivery system suitable for introduction of the antisense sequences into appropriate target cells can be used. Antisense sequences can be delivered intracellularly in the form of an expression plasmid which, upon transcription, produces a sequence complementary to at least a portion of the cellular sequence encoding the target protein. (See, e.g., Slater, J.E. et al. (1998) J. Allergy Clin. Immunol. 102(3):469-475; and Scanlon, K.J. et al. (1995)

9(13):1288-1296.) Antisense sequences can also be introduced intracellularly through the use of viral vectors, such as retrovirus and adeno-associated virus vectors. (See, e.g., Miller, A.D. (1990) Blood 76:271; Ausubel, supra; Uckert, W. and W. Walther (1994) Pharmacol. Ther. 63(3):323-347.) Other gene delivery mechanisms include liposome-derived systems, artificial viral envelopes, and other systems known in the art. (See, e.g., Rossi, J.J. (1995) Br. Med. Bull. 51(1):217-225; Boado, R.J. et al. (1998) J. Pharm. Sci. 87(11):1308-1315; and Morris, M.C. et al. (1997) Nucleic Acids Res. 25(14):2730-2736.)

In another embodiment of the invention, polynucleotides encoding PP may be used for somatic or germline gene therapy. Gene therapy may be performed to (i) correct a genetic deficiency (e.g., in the cases of severe combined immunodeficiency (SCID)-X1 disease characterized by X-linked inheritance (Cavazzana-Calvo, M. et al. (2000) Science 288:669-672), severe combined immunodeficiency syndrome associated with an inherited adenosine deaminase (ADA) deficiency (Blaese, R.M. et al. (1995) Science 270:475-480; Bordignon, C. et al. (1995) Science 270:470-475), cystic fibrosis (Zabner, J. et al. (1993) Cell 75:207-216; Crystal, R.G. et al. (1995) Hum. Gene Therapy 6:643-666; Crystal, R.G. et al. (1995) Hum. Gene Therapy 6:667-703), thalassamias, familial hypercholesterolemia, and hemophilia resulting from Factor VIII or Factor IX deficiencies (Crystal, R.G. (1995) Science 270:404-410; Verma, I.M. and N. Somia (1997) Nature 389:239-242)), (ii) express a conditionally lethal gene product (e.g., in the case of cancers which result from unregulated cell proliferation), or (iii) express a protein which affords protection against intracellular parasites (e.g., against human retroviruses, such as human immunodeficiency virus (HIV) (Baltimore, D. (1988) Nature 335:395-396; Poeschla, E. et al. (1996) Proc. Natl. Acad. Sci. USA. 93:11395-11399), hepatitis B or C virus (HBV, HCV); fungal parasites, such as Candida albicans and Paracoccidioides brasiliensis; and protozoan parasites such as Plasmodium falciparum and Trypanosoma cruzi). In the case where a genetic deficiency in PP expression or regulation causes disease, the expression of PP from an appropriate population of transduced cells may alleviate the clinical manifestations caused by the genetic deficiency.

In a further embodiment of the invention, diseases or disorders caused by deficiencies in PP are treated by constructing mammalian expression vectors encoding PP and introducing these vectors by mechanical means into PP-deficient cells. Mechanical transfer technologies for use with cells in vivo or ex vitro include (i) direct DNA microinjection into individual cells, (ii) ballistic gold particle delivery, (iii) liposome-mediated transfection, (iv) receptor-mediated gene transfer, and (v) the use of DNA transposons (Morgan, R.A. and W.F. Anderson (1993) Annu. Rev. Biochem. 62:191-217; Ivics, Z. (1997) Cell 91:501-510; Boulay, J-L. and H. Récipon (1998) Curr. Opin. Biotechnol. 9:445-450).

Expression vectors that may be effective for the expression of PP include, but are not limited to, the PCDNA 3.1, EPITAG, PRCCMV2, PREP, PVAX, PCR2-TOPOTA vectors (Invitrogen, Carlsbad CA), PCMV-SCRIPT, PCMV-TAG, PEGSH/PERV (Stratagene, La Jolla CA), and PTET-OFF, PTET-ON, PTRE2, PTRE2-LUC, PTK-HYG (Clontech, Palo Alto CA). PP may be expressed using (i) a constitutively active promoter, (e.g., from cytomegalovirus (CMV), Rous sarcoma virus (RSV), SV40 virus, thymidine kinase (TK), or β-actin genes), (ii) an inducible promoter (e.g., the tetracycline-regulated promoter (Gossen, M. and H. Bujard (1992) Proc. Natl. Acad. Sci. USA 89:5547-5551; Gossen, M. et al. (1995) Science 268:1766-1769; Rossi, F.M.V. and H.M. Blau (1998) Curr. Opin. Biotechnol. 9:451-456), commercially available in the T-REX plasmid (Invitrogen)); the ecdysone-inducible promoter (available in the plasmids PVGRXR and PIND; Invitrogen); the FK506/rapamycin inducible promoter; or the RU486/mifepristone inducible promoter (Rossi, F.M.V. and Blau, H.M. supra)), or (iii) a tissue-specific promoter or the native promoter of the endogenous gene encoding PP from a normal individual.

Commercially available liposome transformation kits (e.g., the PERFECT LIPID TRANSFECTION KIT, available from Invitrogen) allow one with ordinary skill in the art to deliver polynucleotides to target cells in culture and require minimal effort to optimize experimental parameters. In the alternative, transformation is performed using the calcium phosphate method (Graham, F.L. and A.J. Eb (1973) Virology 52:456-467), or by electroporation (Neumann, E. et al. (1982) EMBO J. 1:841-845). The introduction of DNA to primary cells requires modification of these standardized mammalian transfection protocols.

In another embodiment of the invention, diseases or disorders caused by genetic defects with respect to PP expression are treated by constructing a retrovirus vector consisting of (i) the polynucleotide encoding PP under the control of an independent promoter or the retrovirus long terminal repeat (LTR) promoter, (ii) appropriate RNA packaging signals, and (iii) a Rev-responsive element (RRE) along with additional retrovirus *cis*-acting RNA sequences and coding sequences required for efficient vector propagation. Retrovirus vectors (e.g., PFB and PFBNEO) are commercially available (Stratagene) and are based on published data (Riviere, I. et al. (1995) Proc. Natl. Acad. Sci. USA 92:6733-6737), incorporated by reference herein. The vector is propagated in an appropriate vector producing cell line (VPCL) that expresses an envelope gene with a tropism for receptors on the target cells or a promiscuous envelope protein such as VSVg (Armentano, D. et al. (1987) J. Virol. 61:1647-1650; Bender, M.A. et al. (1987) J. Virol. 61:1639-1646; Adam, M.A. and A.D. Miller (1988) J. Virol. 62:3802-3806; Dull, T. et al. (1998) J. Virol. 72:8463-8471; Zufferey, R. et al. (1998) J. Virol. 72:9873-9880). U.S. Patent Number 5,910,434 to Rigg ("Method for obtaining retrovirus packaging cell lines producing high transducing efficiency retroviral supernatant") discloses a method for obtaining retrovirus packaging cell lines and is hereby incorporated by reference. Propagation of retrovirus vectors, transduction of a population of cells (e.g., $CD4^+$ T-cells), and the return of transduced cells to a patient are procedures well known to persons skilled in the art of gene therapy and have been well documented (Ranga, U. et al. (1997) J. Virol. 71:7020-7029; Bauer, G. et al. (1997) Blood 89:2259-2267; Bonyhadi, M.L. (1997) J. Virol. 71:4707-4716; Ranga, U. et al. (1998) Proc. Natl. Acad. Sci. USA 95:1201-1206; Su, L. (1997) Blood 89:2283-2290).

In the alternative, an adenovirus-based gene therapy delivery system is used to deliver polynucleotides encoding PP to cells which have one or more genetic abnormalities with respect to the expression of PP. The construction and packaging of adenovirus-based vectors are well known to those with ordinary skill in the art. Replication defective adenovirus vectors have proven to be versatile for importing genes encoding immunoregulatory proteins into intact islets in the pancreas (Csete, M.E. et al. (1995) Transplantation 27:263-268). Potentially useful adenoviral vectors are described in U.S. Patent Number 5,707,618 to Armentano ("Adenovirus vectors for gene therapy"), hereby incorporated by reference. For adenoviral vectors, see also Antinozzi, P.A. et al. (1999) Annu. Rev. Nutr. 19:511-544 and Verma, I.M. and N. Somia (1997) Nature 18:389:239-242, both incorporated by reference herein.

In another alternative, a herpes-based, gene therapy delivery system is used to deliver polynucleotides encoding PP to target cells which have one or more genetic abnormalities with respect to the expression of PP. The use of herpes simplex virus (HSV)-based vectors may be especially valuable for introducing PP to cells of the central nervous system, for which HSV has a tropism. The construction and packaging of herpes-based vectors are well known to those with ordinary skill in the art. A replication-competent herpes simplex virus (HSV) type 1-based vector has been used to deliver a reporter gene to the eyes of primates (Liu, X. et al. (1999) Exp. Eye Res. 169:385-395). The construction of a HSV-1 virus vector has also been disclosed in detail in U.S. Patent Number 5,804,413 to DeLuca ("Herpes simplex virus strains for gene transfer"), which is hereby incorporated by reference. U.S. Patent Number 5,804,413 teaches the use of recombinant HSV d92 which consists of a genome containing at least one exogenous gene to be transferred to a cell under the control of the appropriate promoter for purposes including human gene therapy. Also taught by this patent are the construction and use of recombinant HSV strains deleted for ICP4, ICP27 and ICP22. For HSV vectors, see also Goins, W.F. et al. (1999) J. Virol. 73:519-532 and Xu, H. et al. (1994) Dev. Biol. 163:152-161, hereby incorporated by reference. The manipulation of cloned herpesvirus sequences, the generation of recombinant virus following the transfection of multiple plasmids containing different segments of the large herpesvirus genomes, the growth and propagation of herpesvirus, and the infection of cells with herpesvirus are techniques well known to those of ordinary skill in the art.

In another alternative, an alphavirus (positive, single-stranded RNA virus) vector is used to deliver polynucleotides encoding PP to target cells. The biology of the prototypic alphavirus, Semliki

Forest Virus (SFV), has been studied extensively and gene transfer vectors have been based on the SFV genome (Garoff, H. and K.-J. Li (1998) Curr. Opin. Biotechnol. 9:464-469). During alphavirus RNA replication, a subgenomic RNA is generated that normally encodes the viral capsid proteins. This subgenomic RNA replicates to higher levels than the full length genomic RNA, resulting in the overproduction of capsid proteins relative to the viral proteins with enzymatic activity (e.g., protease and polymerase). Similarly, inserting the coding sequence for PP into the alphavirus genome in place of the capsid-coding region results in the production of a large number of PP-coding RNAs and the synthesis of high levels of PP in vector transduced cells. While alphavirus infection is typically associated with cell lysis within a few days, the ability to establish a persistent infection in hamster normal kidney cells (BHK-21) with a variant of Sindbis virus (SIN) indicates that the lytic replication of alphaviruses can be altered to suit the needs of the gene therapy application (Dryga, S.A. et al. (1997) Virology 228:74-83). The wide host range of alphaviruses will allow the introduction of PP into a variety of cell types. The specific transduction of a subset of cells in a population may require the sorting of cells prior to transduction. The methods of manipulating infectious cDNA clones of alphaviruses, performing alphavirus cDNA and RNA transfections, and performing alphavirus infections, are well known to those with ordinary skill in the art.

Oligonucleotides derived from the transcription initiation site, e.g., between about positions -10 and +10 from the start site, may also be employed to inhibit gene expression. Similarly, inhibition can be achieved using triple helix base-pairing methodology. Triple helix pairing is useful because it causes inhibition of the ability of the double helix to open sufficiently for the binding of polymerases, transcription factors, or regulatory molecules. Recent therapeutic advances using triplex DNA have been described in the literature. (See, e.g., Gee, J.E. et al. (1994) in Huber, B.E. and B.I. Carr, Molecular and Immunologic Approaches, Futura Publishing, Mt. Kisco NY, pp. 163-177.) A complementary sequence or antisense molecule may also be designed to block translation of mRNA by preventing the transcript from binding to ribosomes.

Ribozymes, enzymatic RNA molecules, may also be used to catalyze the specific cleavage of RNA. The mechanism of ribozyme action involves sequence-specific hybridization of the ribozyme molecule to complementary target RNA, followed by endonucleolytic cleavage. For example, engineered hammerhead motif ribozyme molecules may specifically and efficiently catalyze endonucleolytic cleavage of sequences encoding PP.

Specific ribozyme cleavage sites within any potential RNA target are initially identified by scanning the target molecule for ribozyme cleavage sites, including the following sequences: GUA, GUU, and GUC. Once identified, short RNA sequences of between 15 and 20 ribonucleotides, corresponding to the region of the target gene containing the cleavage site, may be evaluated for secondary structural features which may render the oligonucleotide inoperable. The suitability of candidate targets may also be evaluated by testing accessibility to hybridization with complementary oligonucleotides using ribonuclease protection assays.

Complementary ribonucleic acid molecules and ribozymes of the invention may be prepared by any method known in the art for the synthesis of nucleic acid molecules. These include techniques for chemically synthesizing oligonucleotides such as solid phase phosphoramidite chemical synthesis. Alternatively, RNA molecules may be generated by in vitro and in vivo transcription of DNA sequences encoding PP. Such DNA sequences may be incorporated into a wide variety of vectors with suitable RNA polymerase promoters such as T7 or SP6. Alternatively, these cDNA constructs that synthesize complementary RNA, constitutively or inducibly, can be introduced into cell lines, cells, or tissues.

RNA molecules may be modified to increase intracellular stability and half-life. Possible modifications include, but are not limited to, the addition of flanking sequences at the 5' and/or 3' ends of the molecule, or the use of phosphorothioate or 2' O-methyl rather than phosphodiesterase linkages within the backbone of the molecule. This concept is inherent in the production of PNAs and can be extended in all of these molecules by the inclusion of nontraditional bases such as inosine, queosine, and wybutosine, as well as acetyl-, methyl-, thio-, and similarly modified forms of adenine, cytidine, guanine, thymine, and uridine which are not as easily recognized by endogenous endonucleases.

An additional embodiment of the invention encompasses a method for screening for a compound which is effective in altering expression of a polynucleotide encoding PP. Compounds which may be effective in altering expression of a specific polynucleotide may include, but are not limited to, oligonucleotides, antisense oligonucleotides, triple helix-forming oligonucleotides, transcription factors and other polypeptide transcriptional regulators, and non-macromolecular chemical entities which are capable of interacting with specific polynucleotide sequences. Effective compounds may alter polynucleotide expression by acting as either inhibitors or promoters of polynucleotide expression. Thus, in the treatment of disorders associated with increased PP expression or activity, a compound which specifically inhibits expression of the polynucleotide encoding PP may be therapeutically useful, and in the treatment of disorders associated with decreased PP expression or activity, a compound which specifically promotes expression of the polynucleotide encoding PP may be therapeutically useful.

At least one, and up to a plurality, of test compounds may be screened for effectiveness in altering expression of a specific polynucleotide. A test compound may be obtained by any method commonly known in the art, including chemical modification of a compound known to be effective in altering polynucleotide expression; selection from an existing, commercially-available or proprietary library of naturally-occurring or non-natural chemical compounds; rational design of a compound based on chemical and/or structural properties of the target polynucleotide; and selection from a library of chemical compounds created combinatorially or randomly. A sample comprising a polynucleotide encoding PP is exposed to at least one test compound thus obtained. The sample may comprise, for example, an intact or permeabilized cell, or an in vitro cell-free or reconstituted biochemical system. Alterations in the expression of a polynucleotide encoding PP are assayed by any method commonly known in the art. Typically, the expression of a specific nucleotide is detected by hybridization with a probe having a nucleotide sequence complementary to the sequence of the polynucleotide encoding PP. The amount of hybridization may be quantified, thus forming the basis for a comparison of the expression of the polynucleotide both with and without exposure to one or more test compounds. Detection of a change in the expression of a polynucleotide exposed to a test compound indicates that the test compound is effective in altering the expression of the polynucleotide. A screen for a compound effective in altering expression of a specific polynucleotide can be carried out, for example, using a Schizosaccharomyces pombe gene expression system (Atkins, D. et al. (1999) U.S. Patent No. 5,932,435; Arndt, G.M. et al. (2000) Nucleic Acids Res. 28:E15) or a human cell line such as HeLa cell (Clarke, M.L. et al. (2000) Biochem. Biophys. Res. Commun. 268:8-13). A particular embodiment of the present invention involves screening a combinatorial library of oligonucleotides (such as deoxyribonucleotides, ribonucleotides, peptide nucleic acids, and modified oligonucleotides) for antisense activity against a specific polynucleotide sequence (Bruice, T.W. et al. (1997) U.S. Patent No. 5,686,242; Bruice, T.W. et al. (2000) U.S. Patent No. 6,022,691).

Many methods for introducing vectors into cells or tissues are available and equally suitable for use in vivo, in vitro, and ex vivo. For ex vivo therapy, vectors may be introduced into stem cells taken from the patient and clonally propagated for autologous transplant back into that same patient. Delivery by transfection, by liposome injections, or by polycationic amino polymers may be achieved using methods which are well known in the art. (See, e.g., Goldman, C.K. et al. (1997) Nat. Biotechnol. 15:462-466.)

Any of the therapeutic methods described above may be applied to any subject in need of such therapy, including, for example, mammals such as humans, dogs, cats, cows, horses, rabbits, and monkeys.

An additional embodiment of the invention relates to the administration of a composition which generally comprises an active ingredient formulated with a pharmaceutically acceptable excipient. Excipients may include, for example, sugars, starches, celluloses, gums, and proteins. Various formulations are commonly known and are thoroughly discussed in the latest edition of Remington's Pharmaceutical Sciences (Maack Publishing, Easton PA). Such compositions may consist of PP, antibodies to PP, and mimetics, agonists, antagonists, or inhibitors of PP.

The compositions utilized in this invention may be administered by any number of routes including, but not limited to, oral, intravenous, intramuscular, intra-arterial, intramedullary, intrathecal, intraventricular, pulmonary, transdermal, subcutaneous, intraperitoneal, intranasal, enteral, topical, sublingual, or rectal means.

Compositions for pulmonary administration may be prepared in liquid or dry powder form. These compositions are generally aerosolized immediately prior to inhalation by the patient. In the case of small molecules (e.g. traditional low molecular weight organic drugs), aerosol delivery of fast-acting formulations is well-known in the art. In the case of macromolecules (e.g. larger peptides and proteins), recent developments in the field of pulmonary delivery via the alveolar region of the lung have enabled the practical delivery of drugs such as insulin to blood circulation (see, e.g., Patton, J.S. et al., U.S. Patent No. 5,997,848). Pulmonary delivery has the advantage of administration without needle injection, and obviates the need for potentially toxic penetration enhancers.

Compositions suitable for use in the invention include compositions wherein the active ingredients are contained in an effective amount to achieve the intended purpose. The determination of an effective dose is well within the capability of those skilled in the art.

Specialized forms of compositions may be prepared for direct intracellular delivery of macromolecules comprising PP or fragments thereof. For example, liposome preparations containing a cell-impermeable macromolecule may promote cell fusion and intracellular delivery of the macromolecule. Alternatively, PP or a fragment thereof may be joined to a short cationic N-terminal portion from the HIV Tat-1 protein. Fusion proteins thus generated have been found to transduce into the cells of all tissues, including the brain, in a mouse model system (Schwarze, S.R. et al. (1999) Science 285:1569-1572).

For any compound, the therapeutically effective dose can be estimated initially either in cell culture assays, e.g., of neoplastic cells, or in animal models such as mice, rats, rabbits, dogs, monkeys, or pigs. An animal model may also be used to determine the appropriate concentration range and route of administration. Such information can then be used to determine useful doses and routes for administration in humans.

A therapeutically effective dose refers to that amount of active ingredient, for example PP or fragments thereof, antibodies of PP, and agonists, antagonists or inhibitors of PP, which ameliorates the symptoms or condition. Therapeutic efficacy and toxicity may be determined by standard pharmaceutical procedures in cell cultures or with experimental animals, such as by calculating the ED_{50} (the dose therapeutically effective in 50% of the population) or LD_{50} (the dose lethal to 50% of the population) statistics. The dose ratio of toxic to therapeutic effects is the therapeutic index, which can be expressed as the LD_{50}/ED_{50} ratio. Compositions which exhibit large therapeutic indices are preferred. The data obtained from cell culture assays and animal studies are used to formulate a range of dosage for human use. The dosage contained in such compositions is preferably within a range of circulating concentrations that includes the ED_{50} with little or no toxicity. The dosage varies within this range depending upon the dosage form employed, the sensitivity of the patient, and the route of administration.

The exact dosage will be determined by the practitioner, in light of factors related to the subject requiring treatment. Dosage and administration are adjusted to provide sufficient levels of the active moiety or to maintain the desired effect. Factors which may be taken into account include the severity of the disease state, the general health of the subject, the age, weight, and gender of the subject, time and frequency of administration, drug combination(s), reaction sensitivities, and response to therapy. Long-acting compositions may be administered every 3 to 4 days, every week, or biweekly depending on the half-life and clearance rate of the particular formulation.

Normal dosage amounts may vary from about 0.1 μg to 100,000 μg, up to a total dose of about 1 gram, depending upon the route of administration. Guidance as to particular dosages and methods of delivery is provided in the literature and generally available to practitioners in the art. Those skilled in the art will employ different formulations for nucleotides than for proteins or their inhibitors. Similarly, delivery of polynucleotides or polypeptides will be specific to particular cells, conditions, locations, etc.

DIAGNOSTICS

In another embodiment, antibodies which specifically bind PP may be used for the diagnosis of disorders characterized by expression of PP, or in assays to monitor patients being treated with PP or agonists, antagonists, or inhibitors of PP. Antibodies useful for diagnostic purposes may be prepared in the same manner as described above for therapeutics. Diagnostic assays for PP include methods which utilize the antibody and a label to detect PP in human body fluids or in extracts of cells or tissues. The antibodies may be used with or without modification, and may be labeled by covalent or non-covalent attachment of a reporter molecule. A wide variety of reporter molecules, several of which are described above, are known in the art and may be used.

A variety of protocols for measuring PP, including ELISAs, RIAs, and FACS, are known in the art and provide a basis for diagnosing altered or abnormal levels of PP expression. Normal or standard values for PP expression are established by combining body fluids or cell extracts taken from normal mammalian subjects, for example, human subjects, with antibodies to PP under conditions suitable for complex formation. The amount of standard complex formation may be quantitated by various methods, such as photometric means. Quantities of PP expressed in subject, control, and disease samples from biopsied tissues are compared with the standard values. Deviation between standard and subject values establishes the parameters for diagnosing disease.

In another embodiment of the invention, the polynucleotides encoding PP may be used for diagnostic purposes. The polynucleotides which may be used include oligonucleotide sequences, complementary RNA and DNA molecules, and PNAs. The polynucleotides may be used to detect and quantify gene expression in biopsied tissues in which expression of PP may be correlated with disease. The diagnostic assay may be used to determine absence, presence, and excess expression of PP, and to monitor regulation of PP levels during therapeutic intervention.

In one aspect, hybridization with PCR probes which are capable of detecting polynucleotide sequences, including genomic sequences, encoding PP or closely related molecules may be used to identify nucleic acid sequences which encode PP. The specificity of the probe, whether it is made from a highly specific region, e.g., the 5' regulatory region, or from a less specific region, e.g., a conserved motif, and the stringency of the hybridization or amplification will determine whether the probe identifies only naturally occurring sequences encoding PP, allelic variants, or related sequences.

Probes may also be used for the detection of related sequences, and may have at least 50% sequence identity to any of the PP encoding sequences. The hybridization probes of the subject invention may be DNA or RNA and may be derived from the sequence of SEQ ID NO:13-24 or from genomic sequences including promoters, enhancers, and introns of the PP gene.

Means for producing specific hybridization probes for DNAs encoding PP include the cloning of polynucleotide sequences encoding PP or PP derivatives into vectors for the production of mRNA probes. Such vectors are known in the art, are commercially available, and may be used to synthesize RNA probes in vitro by means of the addition of the appropriate RNA polymerases and the appropriate labeled nucleotides. Hybridization probes may be labeled by a variety of reporter groups, for example, by radionuclides such as ^{32}P or ^{35}S, or by enzymatic labels, such as alkaline phosphatase coupled to the probe via avidin/biotin coupling systems, and the like.

Polynucleotide sequences encoding PP may be used for the diagnosis of disorders associated with expression of PP. Examples of such disorders include, but are not limited to, an immune system disorder, such as acquired immunodeficiency syndrome (AIDS), X-linked agammaglobinemia of Bruton, common variable immunodeficiency (CVI), DiGeorge's syndrome (thymic hypoplasia), thymic dysplasia, isolated IgA deficiency, severe combined immunodeficiency disease (SCID), immunodeficiency with thrombocytopenia and eczema (Wiskott-Aldrich syndrome), Chediak-Higashi syndrome, chronic granulomatous diseases, hereditary angioneurotic edema, immunodeficiency associated with Cushing's disease, Addison's disease, adult respiratory distress syndrome, allergies, ankylosing spondylitis, amyloidosis, anemia, asthma, atherosclerosis, autoimmune hemolytic anemia, autoimmune thyroiditis, autoimmune polyendocrinopathy-candidiasis-ectodermal dystrophy (APECED), bronchitis, cholecystitis, contact dermatitis, Crohn's disease, atopic dermatitis, dermatomyositis, diabetes mellitus, emphysema, episodic lymphopenia with lymphocytotoxins, erythroblastosis fetalis, erythema nodosum, atrophic gastritis, glomerulonephritis, Goodpasture's syndrome, gout, Graves' disease, Hashimoto's thyroiditis, hypereosinophilia, irritable bowel syndrome, multiple sclerosis, myasthenia gravis, myocardial or pericardial inflammation, osteoarthritis, osteoporosis, pancreatitis, polymyositis, psoriasis, Reiter's syndrome, rheumatoid arthritis, scleroderma, Sjögren's syndrome, systemic anaphylaxis, systemic lupus erythematosus, systemic sclerosis, thrombocytopenic purpura, ulcerative colitis, uveitis, Werner syndrome, complications of cancer, hemodialysis, and extracorporeal circulation, viral, bacterial, fungal, parasitic, protozoal, and helminthic infections, and trauma; a neurological disorder, such as epilepsy, ischemic cerebrovascular disease, stroke, cerebral neoplasms, Alzheimer's disease, Pick's disease, Huntington's disease, dementia, Parkinson's disease and other extrapyramidal disorders, amyotrophic lateral sclerosis and other motor neuron disorders, progressive neural muscular atrophy, retinitis pigmentosa, hereditary ataxias, multiple sclerosis and other demyelinating diseases, bacterial and viral meningitis, brain abscess, subdural empyema, epidural abscess, suppurative intracranial thrombophlebitis, myelitis and radiculitis, viral central nervous system disease, prion diseases including kuru, Creutzfeldt-Jakob disease, and Gerstmann-Straussler-Scheinker syndrome, fatal familial insomnia, nutritional and metabolic diseases of the nervous system, neurofibromatosis, tuberous sclerosis, cerebelloretinal hemangioblastomatosis, encephalotrigeminal syndrome, mental retardation and other developmental disorders of the central nervous system including Down syndrome, cerebral palsy, neuroskeletal disorders, autonomic nervous system disorders, cranial nerve disorders, spinal cord diseases, muscular dystrophy and other neuromuscular disorders, peripheral nervous system disorders, dermatomyositis and polymyositis, inherited, metabolic, endocrine, and toxic myopathies, myasthenia gravis, periodic paralysis, mental disorders including mood, anxiety, and schizophrenic disorders, seasonal affective disorder (SAD), akathesia, amnesia, catatonia, diabetic neuropathy, tardive dyskinesia, dystonias, paranoid psychoses, postherpetic neuralgia, Tourette's disorder, progressive supranuclear palsy, corticobasal degeneration, and familial frontotemporal dementia; a developmental disorder, such as renal tubular acidosis, anemia, Cushing's syndrome, achondroplastic dwarfism, Duchenne and Becker muscular dystrophy, epilepsy, gonadal dysgenesis, WAGR syndrome (Wilms' tumor, aniridia, genitourinary abnormalities, and mental retardation), Smith-Magenis syndrome, myelodysplastic syndrome, hereditary mucoepithelial dysplasia, hereditary keratodermas, hereditary neuropathies such as Charcot-Marie-Tooth disease and neurofibromatosis, hypothyroidism, hydrocephalus, seizure disorders such as Syndenham's chorea and cerebral palsy, spina bifida, anencephaly, craniorachischisis, congenital glaucoma, cataract, and sensorineural hearing loss; and a cell proliferative disorder, such as actinic keratosis, arteriosclerosis, atherosclerosis, bursitis, cirrhosis, hepatitis, mixed connective tissue disease (MCTD), myelofibrosis, paroxysmal nocturnal hemoglobinuria, polycythemia vera, psoriasis, primary thrombocythemia, and cancers including adenocarcinoma, leukemia, lymphoma, melanoma, myeloma, sarcoma, teratocarcinoma, and, in particular, cancers of the adrenal gland, bladder, bone, bone marrow, brain, breast, cervix, gall bladder, ganglia, gastrointestinal tract, heart, kidney, liver, lung, muscle, ovary, pancreas, parathyroid, penis, prostate, salivary glands, skin, spleen, testis, thymus, thyroid, and uterus. The polynucleotide sequences encoding PP may be used in Southern or northern analysis, dot blot, or other membrane-based technologies; in PCR technologies; in dipstick, pin, and multiformat ELISA-like assays; and in microarrays utilizing fluids or tissues from patients to detect altered PP expression. Such qualitative or quantitative methods are well known in the art.

In a particular aspect, the nucleotide sequences encoding PP may be useful in assays that detect the presence of associated disorders, particularly those mentioned above. The nucleotide sequences encoding PP may be labeled by standard methods and added to a fluid or tissue sample from a patient under conditions suitable for the formation of hybridization complexes. After a suitable incubation period, the sample is washed and the signal is quantified and compared with a standard value. If the amount of signal in the patient sample is significantly altered in comparison to a control sample then the presence of altered levels of nucleotide sequences encoding PP in the sample indicates the presence of the associated disorder. Such assays may also be used to evaluate the efficacy of a particular therapeutic treatment regimen in animal studies, in clinical trials, or to monitor the treatment of an individual patient.

In order to provide a basis for the diagnosis of a disorder associated with expression of PP, a normal or standard profile for expression is established. This may be accomplished by combining body fluids or cell extracts taken from normal subjects, either animal or human, with a sequence, or a fragment thereof, encoding PP, under conditions suitable for hybridization or amplification. Standard hybridization may be quantified by comparing the values obtained from normal subjects with values from an experiment in which a known amount of a substantially purified polynucleotide is used. Standard values obtained in this manner may be compared with values obtained from samples from patients who are symptomatic for a disorder. Deviation from standard values is used to establish the presence of a disorder.

Once the presence of a disorder is established and a treatment protocol is initiated, hybridization assays may be repeated on a regular basis to determine if the level of expression in the patient begins to approximate that which is observed in the normal subject. The results obtained from successive assays may be used to show the efficacy of treatment over a period ranging from several days to months.

With respect to cancer, the presence of an abnormal amount of transcript (either under- or overexpressed) in biopsied tissue from an individual may indicate a predisposition for the development of the disease, or may provide a means for detecting the disease prior to the appearance of actual clinical symptoms. A more definitive diagnosis of this type may allow health professionals to employ preventative measures or aggressive treatment earlier thereby preventing the development or further progression of the cancer.

Additional diagnostic uses for oligonucleotides designed from the sequences encoding PP may involve the use of PCR. These oligomers may be chemically synthesized, generated enzymatically, or produced in vitro. Oligomers will preferably contain a fragment of a polynucleotide encoding PP, or a fragment of a polynucleotide complementary to the polynucleotide encoding PP, and will be employed under optimized conditions for identification of a specific gene or condition. Oligomers may also be employed under less stringent conditions for detection or quantification of closely related DNA or RNA sequences.

In a particular aspect, oligonucleotide primers derived from the polynucleotide sequences encoding PP may be used to detect single nucleotide polymorphisms (SNPs). SNPs are substitutions, insertions and deletions that are a frequent cause of inherited or acquired genetic disease in humans. Methods of SNP detection include, but are not limited to, single-stranded conformation polymorphism (SSCP) and fluorescent SSCP (fSSCP) methods. In SSCP, oligonucleotide primers derived from the polynucleotide sequences encoding PP are used to amplify DNA using the polymerase chain reaction (PCR). The DNA may be derived, for example, from diseased or normal tissue, biopsy samples, bodily fluids, and the like. SNPs in the DNA cause differences in the secondary and tertiary structures of PCR products in single-stranded form, and these differences are detectable using gel electrophoresis in non-denaturing gels. In fSCCP, the oligonucleotide primers are fluorescently labeled, which allows detection of the amplimers in high-throughput equipment such as DNA sequencing machines. Additionally, sequence database analysis methods, termed in silico SNP (isSNP), are capable of identifying polymorphisms by comparing the sequence of individual overlapping DNA fragments which assemble into a common consensus sequence. These computer-based methods filter out sequence variations due to laboratory preparation of DNA and sequencing errors using statistical models and automated analyses of DNA sequence chromatograms. In the alternative, SNPs may be detected and characterized by mass spectrometry using, for example, the high throughput MASSARRAY system (Sequenom, Inc., San Diego CA).

Methods which may also be used to quantify the expression of PP include radiolabeling or biotinylating nucleotides, coamplification of a control nucleic acid, and interpolating results from standard curves. (See, e.g., Melby, P.C. et al. (1993) J. Immunol. Methods 159:235-244; Duplaa, C. et al. (1993) Anal. Biochem. 212:229-236.) The speed of quantitation of multiple samples may be accelerated by running the assay in a high-throughput format where the oligomer or polynucleotide of interest is presented in various dilutions and a spectrophotometric or colorimetric response gives rapid quantitation.

In further embodiments, oligonucleotides or longer fragments derived from any of the polynucleotide sequences described herein may be used as elements on a microarray. The microarray can be used in transcript imaging techniques which monitor the relative expression levels of large numbers of genes simultaneously as described below. The microarray may also be used to identify genetic variants, mutations, and polymorphisms. This information may be used to determine gene function, to understand the genetic basis of a disorder, to diagnose a disorder, to monitor progression/regression of disease as a function of gene expression, and to develop and monitor the activities of therapeutic agents in the treatment of disease. In particular, this information may be used to develop a pharmacogenomic profile of a patient in order to select the most appropriate and effective treatment regimen for that patient. For example, therapeutic agents which are highly effective and display the fewest side effects may be selected for a patient based on his/her pharmacogenomic profile.

In another embodiment, PP, fragments of PP, or antibodies specific for PP may be used as elements on a microarray. The microarray may be used to monitor or measure protein-protein interactions, drug-target interactions, and gene expression profiles, as described above.

A particular embodiment relates to the use of the polynucleotides of the present invention to generate a transcript image of a tissue or cell type. A transcript image represents the global pattern of gene expression by a particular tissue or cell type. Global gene expression patterns are analyzed by quantifying the number of expressed genes and their relative abundance under given conditions and at a given time. (See Seilhamer et al., "Comparative Gene Transcript Analysis," U.S. Patent Number 5,840,484, expressly incorporated by reference herein.) Thus a transcript image may be generated by hybridizing the polynucleotides of the present invention or their complements to the totality of transcripts or reverse transcripts of a particular tissue or cell type. In one embodiment, the hybridization takes place in high-throughput format, wherein the polynucleotides of the present invention or their complements comprise a subset of a plurality of elements on a microarray. The resultant transcript image would provide a profile of gene activity.

Transcript images may be generated using transcripts isolated from tissues, cell lines, biopsies, or other biological samples. The transcript image may thus reflect gene expression in vivo, as in the case of a tissue or biopsy sample, or in vitro, as in the case of a cell line.

Transcript images which profile the expression of the polynucleotides of the present invention may also be used in conjunction with in vitro model systems and preclinical evaluation of pharmaceuticals, as well as toxicological testing of industrial and naturally-occurring environmental compounds. All compounds induce characteristic gene expression patterns, frequently termed molecular fingerprints or toxicant signatures, which are indicative of mechanisms of action and toxicity (Nuwaysir, E.F. et al. (1999) Mol. Carcinog. 24:153-159; Steiner, S. and N.L. Anderson (2000) Toxicol. Lett. 112-113:467-471, expressly incorporated by reference herein). If a test compound has a signature similar to that of a compound with known toxicity, it is likely to share those toxic properties. These fingerprints or signatures are most useful and refined when they contain expression information from a large number of genes and gene families. Ideally, a genome-wide measurement of expression provides the highest quality signature. Even genes whose expression is not altered by any tested compounds are important as well, as the levels of expression of these genes are used to normalize the rest of the expression data. The normalization procedure is useful for comparison of expression data after treatment with different compounds. While the assignment of gene function to elements of a toxicant signature aids in interpretation of toxicity mechanisms, knowledge of gene function is not necessary for the statistical matching of signatures which leads to prediction of toxicity. (See, for example, Press Release 00-02 from the National Institute of Environmental Health Sciences, released February 29, 2000, available at http://www.niehs.nih.gov/oc/news/toxchip.htm.) Therefore, it is important and desirable in toxicological screening using toxicant signatures to include all expressed gene sequences.

In one embodiment, the toxicity of a test compound is assessed by treating a biological sample containing nucleic acids with the test compound. Nucleic acids that are expressed in the treated biological sample are hybridized with one or more probes specific to the polynucleotides of the present invention, so that transcript levels corresponding to the polynucleotides of the present invention may be quantified. The transcript levels in the treated biological sample are compared with levels in an untreated biological sample. Differences in the transcript levels between the two samples are indicative of a toxic response caused by the test compound in the treated sample.

Another particular embodiment relates to the use of the polypeptide sequences of the present invention to analyze the proteome of a tissue or cell type. The term proteome refers to the global pattern of protein expression in a particular tissue or cell type. Each protein component of a proteome can be subjected individually to further analysis. Proteome expression patterns, or profiles, are analyzed by quantifying the number of expressed proteins and their relative abundance under given conditions and at a given time. A profile of a cell's proteome may thus be generated by separating and analyzing the polypeptides of a particular tissue or cell type. In one embodiment, the separation is achieved using two-dimensional gel electrophoresis, in which proteins from a sample are separated by isoelectric focusing in the first dimension, and then according to molecular weight by sodium dodecyl sulfate slab gel electrophoresis in the second dimension (Steiner and Anderson, supra). The proteins are visualized in the gel as discrete and uniquely positioned spots, typically by staining the gel with an agent such as Coomassie Blue or silver or fluorescent stains. The optical density of each protein spot is generally proportional to the level of the protein in the sample. The optical densities of equivalently positioned protein spots from different samples, for example, from biological samples either treated or untreated with a test compound or therapeutic agent, are compared to identify any changes in protein spot density related to the treatment. The proteins in the spots are partially sequenced using, for example, standard methods employing chemical or enzymatic cleavage followed by mass spectrometry. The identity of the protein in a spot may be determined by comparing its partial sequence, preferably of at least 5 contiguous amino acid residues, to the polypeptide sequences of the present invention. In some cases, further sequence data may be obtained for definitive protein identification.

A proteomic profile may also be generated using antibodies specific for PP to quantify the levels of PP expression. In one embodiment, the antibodies are used as elements on a microarray, and protein expression levels are quantified by exposing the microarray to the sample and detecting the levels of protein bound to each array element (Lueking, A. et al. (1999) Anal. Biochem. 270:103-111; Mendoze, L.G. et al. (1999) Biotechniques 27:778-788). Detection may be performed by a variety of methods known in the art, for example, by reacting the proteins in the sample with a thiol- or amino-reactive fluorescent compound and detecting the amount of fluorescence bound at each array element.

Toxicant signatures at the proteome level are also useful for toxicological screening, and should be analyzed in parallel with toxicant signatures at the transcript level. There is a poor correlation between transcript and protein abundances for some proteins in some tissues (Anderson, N.L. and J. Seilhamer (1997) Electrophoresis 18:533-537), so proteome toxicant signatures may be useful in the analysis of compounds which do not significantly affect the transcript image, but which alter the proteomic profile. In addition, the analysis of transcripts in body fluids is difficult, due to rapid degradation of mRNA, so proteomic profiling may be more reliable and informative in such cases.

In another embodiment, the toxicity of a test compound is assessed by treating a biological sample containing proteins with the test compound. Proteins that are expressed in the treated biological sample are separated so that the amount of each protein can be quantified. The amount of each protein is compared to the amount of the corresponding protein in an untreated biological sample. A difference in the amount of protein between the two samples is indicative of a toxic response to the test compound in the treated sample. Individual proteins are identified by sequencing the amino acid residues of the individual proteins and comparing these partial sequences to the polypeptides of the present invention.

In another embodiment, the toxicity of a test compound is assessed by treating a biological sample containing proteins with the test compound. Proteins from the biological sample are incubated with antibodies specific to the polypeptides of the present invention. The amount of protein recognized by the antibodies is quantified. The amount of protein in the treated biological sample is compared with the amount in an untreated biological sample. A difference in the amount of protein between the two samples is indicative of a toxic response to the test compound in the treated sample.

Microarrays may be prepared, used, and analyzed using methods known in the art. (See, e.g., Brennan, T.M. et al. (1995) U.S. Patent No. 5,474,796; Schena, M. et al. (1996) Proc. Natl. Acad. Sci. USA 93:10614-10619; Baldeschweiler et al. (1995) PCT application WO95/251116; Shalon, D. et al. (1995) PCT application WO95/35505; Heller, R.A. et al. (1997) Proc. Natl. Acad. Sci. USA 94:2150-2155; and Heller, M.J. et al. (1997) U.S. Patent No. 5,605,662.) Various types of microarrays are well known and thoroughly described in DNA Microarrays: A Practical Approach, M. Schena, ed. (1999) Oxford University Press, London, hereby expressly incorporated by reference.

In another embodiment of the invention, nucleic acid sequences encoding PP may be used to generate hybridization probes useful in mapping the naturally occurring genomic sequence. Either coding or noncoding sequences may be used, and in some instances, noncoding sequences may be preferable over coding sequences. For example, conservation of a coding sequence among members of a multi-gene family may potentially cause undesired cross hybridization during chromosomal mapping. The sequences may be mapped to a particular chromosome, to a specific region of a chromosome, or to artificial chromosome constructions, e.g., human artificial chromosomes (HACs), yeast artificial chromosomes (YACs), bacterial artificial chromosomes (BACs), bacterial P1 constructions, or single chromosome cDNA libraries. (See, e.g., Harrington, J.J. et al. (1997) Nat. Genet. 15:345-355; Price, C.M. (1993) Blood Rev. 7:127-134; and Trask, B.J. (1991) Trends Genet. 7:149-154.) Once mapped, the nucleic acid sequences of the invention may be used to develop genetic linkage maps, for example, which correlate the inheritance of a disease state with the inheritance of a particular chromosome region or restriction fragment length polymorphism (RFLP). (See, for example, Lander, E.S. and D. Botstein (1986) Proc. Natl. Acad. Sci. USA 83:7353-7357.)

Fluorescent in situ hybridization (FISH) may be correlated with other physical and genetic map data. (See, e.g., Heinz-Ulrich, et al. (1995) in Meyers, supra, pp. 965-968.) Examples of genetic map data can be found in various scientific journals or at the Online Mendelian Inheritance in Man (OMIM) World Wide Web site. Correlation between the location of the gene encoding PP on a physical map and a specific disorder, or a predisposition to a specific disorder, may help define the region of DNA associated with that disorder and thus may further positional cloning efforts.

In situ hybridization of chromosomal preparations and physical mapping techniques, such as linkage analysis using established chromosomal markers, may be used for extending genetic maps. Often the placement of a gene on the chromosome of another mammalian species, such as mouse, may reveal associated markers even if the exact chromosomal locus is not known. This information is valuable to investigators searching for disease genes using positional cloning or other gene discovery techniques. Once the gene or genes responsible for a disease or syndrome have been crudely localized by genetic linkage to a particular genomic region, e.g., ataxia-telangiectasia to 11q22-23, any sequences mapping to that area may represent associated or regulatory genes for further investigation. (See, e.g., Gatti, R.A. et al. (1988) Nature 336:577-580.) The nucleotide sequence of the instant invention may also be used to detect differences in the chromosomal location due to translocation, inversion, etc., among normal, carrier, or affected individuals.

In another embodiment of the invention, PP, its catalytic or immunogenic fragments, or oligopeptides thereof can be used for screening libraries of compounds in any of a variety of drug screening techniques. The fragment employed in such screening may be free in solution, affixed to a solid support, borne on a cell surface, or located intracellularly. The formation of binding complexes between PP and the agent being tested may be measured.

Another technique for drug screening provides for high throughput screening of compounds having suitable binding affinity to the protein of interest. (See, e.g., Geysen, et al. (1984) PCT application WO84/03564.) In this method, large numbers of different small test compounds are synthesized on a solid substrate. The test compounds are reacted with PP, or fragments thereof, and washed. Bound PP is then detected by methods well known in the art. Purified PP can also be coated directly onto plates for use in the aforementioned drug screening techniques. Alternatively, non-neutralizing antibodies can be used to capture the peptide and immobilize it on a solid support.

In another embodiment, one may use competitive drug screening assays in which neutralizing antibodies capable of binding PP specifically compete with a test compound for binding PP. In this manner, antibodies can be used to detect the presence of any peptide which shares one or more antigenic determinants with PP.

In additional embodiments, the nucleotide sequences which encode PP may be used in any molecular biology techniques that have yet to be developed, provided the new techniques rely on properties of nucleotide sequences that are currently known, including, but not limited to, such properties as the triplet genetic code and specific base pair interactions.

Without further elaboration, it is believed that one skilled in the art can, using the preceding description, utilize the present invention to its fullest extent. The following embodiments are, therefore, to be construed as merely illustrative, and not limitative of the remainder of the disclosure in any way whatsoever.

The disclosures of all patents, applications and publications, mentioned above and below, including U.S. Ser. No. 60/234,526, U.S. Ser. No. 60/236,967, U.S. Ser No. 60/238,332, U.S. Ser. No. 60/242,236, U.S. Ser. No. 60/243,928 and U.S. Ser. No. 60/249,814, are expressly incorporated by reference herein.

EXAMPLES

I. Construction of cDNA Libraries

Incyte cDNAs were derived from cDNA libraries described in the LIFESEQ GOLD database (Incyte Genomics, Palo Alto CA) and shown in Table 4, column 5. Some tissues were homogenized and lysed in guanidinium isothiocyanate, while others were homogenized and lysed in phenol or in a suitable mixture of denaturants, such as TRIZOL (Life Technologies), a monophasic solution of phenol and guanidine isothiocyanate. The resulting lysates were centrifuged over CsCl cushions or extracted with chloroform. RNA was precipitated from the lysates with either isopropanol or sodium acetate and ethanol, or by other routine methods.

Phenol extraction and precipitation of RNA were repeated as necessary to increase RNA purity. In some cases, RNA was treated with DNase. For most libraries, poly(A)+ RNA was isolated using oligo d(T)-coupled paramagnetic particles (Promega), OLIGOTEX latex particles (QIAGEN, Chatsworth CA), or an OLIGOTEX mRNA purification kit (QIAGEN). Alternatively, RNA was isolated directly from tissue lysates using other RNA isolation kits, e.g., the POLY(A)PURE mRNA purification kit (Ambion, Austin TX).

In some cases, Stratagene was provided with RNA and constructed the corresponding cDNA libraries. Otherwise, cDNA was synthesized and cDNA libraries were constructed with the UNIZAP vector system (Stratagene) or SUPERSCRIPT plasmid system (Life Technologies), using the recommended procedures or similar methods known in the art. (See, e.g., Ausubel, 1997, supra, units 5.1-6.6.) Reverse transcription was initiated using oligo d(T) or random primers. Synthetic oligonucleotide adapters were ligated to double stranded cDNA, and the cDNA was digested with the appropriate restriction enzyme or enzymes. For most libraries, the cDNA was size-selected (300-1000 bp) using SEPHACRYL S1000, SEPHAROSE CL2B, or SEPHAROSE CL4B column chromatography (Amersham Pharmacia Biotech) or preparative agarose gel electrophoresis. cDNAs were ligated into compatible restriction enzyme sites of the polylinker of a suitable plasmid, e.g., PBLUESCRIPT plasmid (Stratagene), PSPORT1 plasmid (Life Technologies), PCDNA2.1 plasmid (Invitrogen, Carlsbad CA), PBK-CMV plasmid (Stratagene), PCR2-TOPOTA (Invitrogen), PCMV-ICIS (Stratagene), or pINCY (Incyte Genomics, Palo Alto CA), or derivatives thereof. Recombinant plasmids were transformed into competent E. coli cells including XL1-Blue, XL1-BlueMRF, or SOLR from Stratagene or DH5α, DH10B, or ElectroMAX DH10B from Life Technologies.

II. Isolation of cDNA Clones

Plasmids obtained as described in Example I were recovered from host cells by in vivo excision using the UNIZAP vector system (Stratagene) or by cell lysis. Plasmids were purified using at least one of the following: a Magic or WIZARD Minipreps DNA purification system (Promega); an AGTC Miniprep purification kit (Edge Biosystems, Gaithersburg MD); and QIAWELL 8 Plasmid, QIAWELL 8 Plus Plasmid, QIAWELL 8 Ultra Plasmid purification systems or the R.E.A.L. PREP 96 plasmid purification kit from QIAGEN. Following precipitation, plasmids were resuspended in 0.1 ml of distilled water and stored, with or without lyophilization, at 4°C.

Alternatively, plasmid DNA was amplified from host cell lysates using direct link PCR in a high-throughput format (Rao, V.B. (1994) Anal. Biochem. 216:1-14). Host cell lysis and thermal cycling steps were carried out in a single reaction mixture. Samples were processed and stored in 384-well plates, and the concentration of amplified plasmid DNA was quantified fluorometrically using PICOGREEN dye (Molecular Probes, Eugene OR) and a FLUOROSKAN II fluorescence scanner (Labsystems Oy, Helsinki, Finland).

III. Sequencing and Analysis

Incyte cDNA recovered in plasmids as described in Example II were sequenced as follows. Sequencing reactions were processed using standard methods or high-throughput instrumentation such as the ABI CATALYST 800 (Applied Biosystems) thermal cycler or the PTC-200 thermal cycler (MJ Research) in conjunction with the HYDRA microdispenser (Robbins Scientific) or the MICROLAB 2200 (Hamilton) liquid transfer system. cDNA sequencing reactions were prepared using reagents provided by Amersham Pharmacia Biotech or supplied in ABI sequencing kits such as the ABI PRISM BIGDYE Terminator cycle sequencing ready reaction kit (Applied Biosystems). Electrophoretic separation of cDNA sequencing reactions and detection of labeled polynucleotides were carried out using the MEGABACE 1000 DNA sequencing system (Molecular Dynamics); the ABI PRISM 373 or 377 sequencing system (Applied Biosystems) in conjunction with standard ABI protocols and base calling software; or other sequence analysis systems known in the art. Reading frames within the cDNA sequences were identified using standard methods (reviewed in Ausubel, 1997, supra, unit 7.7). Some of the cDNA sequences were selected for extension using the techniques disclosed in Example VIII.

The polynucleotide sequences derived from Incyte cDNAs were validated by removing vector, linker, and poly(A) sequences and by masking ambiguous bases, using algorithms and programs based on BLAST, dynamic programming, and dinucleotide nearest neighbor analysis. The Incyte cDNA sequences or translations thereof were then queried against a selection of public databases such as the GenBank primate, rodent, mammalian, vertebrate, and eukaryote databases, and BLOCKS, PRINTS, DOMO, PRODOM, and hidden Markov model (HMM)-based protein family databases such as PFAM. (HMM is a probabilistic approach which analyzes consensus primary structures of gene families. See, for example, Eddy, S.R. (1996) Curr. Opin. Struct. Biol. 6:361-365.) The queries were performed using programs based on BLAST, FASTA, BLIMPS, and HMMER. The Incyte cDNA sequences were assembled to produce full length polynucleotide sequences. Alternatively, GenBank cDNAs, GenBank ESTs, stitched sequences, stretched sequences, or

Genscan-predicted coding sequences (see Examples IV and V) were used to extend Incyte cDNA assemblages to full length. Assembly was performed using programs based on Phred, Phrap, and Consed, and cDNA assemblages were screened for open reading frames using programs based on GeneMark, BLAST, and FASTA. The full length polynucleotide sequences were translated to derive the corresponding full length polypeptide sequences. Alternatively, a polypeptide of the invention may begin at any of the methionine residues of the full length translated polypeptide. Full length polypeptide sequences were subsequently analyzed by querying against databases such as the GenBank protein databases (genpept), SwissProt, BLOCKS, PRINTS, DOMO, PRODOM, Prosite, and hidden Markov model (HMM)-based protein family databases such as PFAM. Full length polynucleotide sequences are also analyzed using MACDNASIS PRO software (Hitachi Software Engineering, South San Francisco CA) and LASERGENE software (DNASTAR). Polynucleotide and polypeptide sequence alignments are generated using default parameters specified by the CLUSTAL algorithm as incorporated into the MEGALIGN multisequence alignment program (DNASTAR), which also calculates the percent identity between aligned sequences.

Table 7 summarizes the tools, programs, and algorithms used for the analysis and assembly of Incyte cDNA and full length sequences and provides applicable descriptions, references, and threshold parameters. The first column of Table 7 shows the tools, programs, and algorithms used, the second column provides brief descriptions thereof, the third column presents appropriate references, all of which are incorporated by reference herein in their entirety, and the fourth column presents, where applicable, the scores, probability values, and other parameters used to evaluate the strength of a match between two sequences (the higher the score or the lower the probability value, the greater the identity between two sequences).

The programs described above for the assembly and analysis of full length polynucleotide and polypeptide sequences were also used to identify polynucleotide sequence fragments from SEQ ID NO:13-24. Fragments from about 20 to about 4000 nucleotides which are useful in hybridization and amplification technologies are described in Table 4, column 4.

IV. Identification and Editing of Coding Sequences from Genomic DNA

Putative protein phosphatases were initially identified by running the Genscan gene identification program against public genomic sequence databases (e.g., gbpri and gbhtg). Genscan is a general-purpose gene identification program which analyzes genomic DNA sequences from a variety of organisms (See Burge, C. and S. Karlin (1997) J. Mol. Biol. 268:78-94, and Burge, C. and S. Karlin (1998) Curr. Opin. Struct. Biol. 8:346-354). The program concatenates predicted exons to form an assembled cDNA sequence extending from a methionine to a stop codon. The output of Genscan is a FASTA database of polynucleotide and polypeptide sequences. The maximum range of sequence for Genscan to analyze at once was set to 30 kb. To determine which of these Genscan predicted cDNA sequences encode protein phosphatases, the encoded polypeptides were analyzed by querying against PFAM models for protein phosphatases. Potential protein phosphatases were also identified by homology to Incyte cDNA sequences that had been annotated as protein phosphatases. These selected Genscan-predicted sequences were then compared by BLAST analysis to the genpept and gbpri public databases. Where necessary, the Genscan-predicted sequences were then edited by comparison to the top BLAST hit from genpept to correct errors in the sequence predicted by Genscan, such as extra or omitted exons. BLAST analysis was also used to find any Incyte cDNA or public cDNA coverage of the Genscan-predicted sequences, thus providing evidence for transcription. When Incyte cDNA coverage was available, this information was used to correct or confirm the Genscan predicted sequence. Full length polynucleotide sequences were obtained by assembling Genscan-predicted coding sequences with Incyte cDNA sequences and/or public cDNA sequences using the assembly process described in Example III. Alternatively, full length polynucleotide sequences were derived entirely from edited or unedited Genscan-predicted coding sequences.

V. Assembly of Genomic Sequence Data with cDNA Sequence Data

"Stitched" Sequences

Partial cDNA sequences were extended with exons predicted by the Genscan gene identification program described in Example IV. Partial cDNAs assembled as described in Example III were mapped to genomic DNA and parsed into clusters containing related cDNAs and Genscan exon predictions from one or more genomic sequences. Each cluster was analyzed using an algorithm based on graph theory and dynamic programming to integrate cDNA and genomic information, generating possible splice variants that were subsequently confirmed, edited, or extended to create a full length sequence. Sequence intervals in which the entire length of the interval was present on more than one sequence in the cluster were identified, and intervals thus identified were considered to be equivalent by transitivity. For example, if an interval was present on a cDNA and two genomic sequences, then all three intervals were considered to be equivalent. This process allows unrelated but consecutive genomic sequences to be brought together, bridged by cDNA sequence. Intervals thus identified were then "stitched" together by the stitching algorithm in the order that they appear along their parent sequences to generate the longest possible sequence, as well as sequence variants. Linkages between intervals which proceed along one type of parent sequence (cDNA to cDNA or genomic sequence to genomic sequence) were given preference over linkages which change parent type (cDNA to genomic sequence). The resultant stitched sequences were translated and compared by BLAST analysis to the genpept and gbpri public databases. Incorrect exons predicted by Genscan were corrected by comparison to the top BLAST hit from genpept. Sequences were further extended with additional cDNA sequences, or by inspection of genomic DNA, when necessary.

"Stretched" Sequences

Partial DNA sequences were extended to full length with an algorithm based on BLAST analysis. First, partial cDNAs assembled as described in Example III were queried against public databases such as the GenBank primate, rodent, mammalian, vertebrate, and eukaryote databases using the BLAST program. The nearest GenBank protein homolog was then compared by BLAST analysis to either Incyte cDNA sequences or GenScan exon predicted sequences described in Example IV. A chimeric protein was generated by using the resultant high-scoring segment pairs (HSPs) to map the translated sequences onto the GenBank protein homolog. Insertions or deletions may occur in the chimeric protein with respect to the original GenBank protein homolog. The GenBank protein homolog, the chimeric protein, or both were used as probes to search for homologous genomic sequences from the public human genome databases. Partial DNA sequences were therefore "stretched" or extended by the addition of homologous genomic sequences. The resultant stretched sequences were examined to determine whether it contained a complete gene.

VI. Chromosomal Mapping of PP Encoding Polynucleotides

The sequences which were used to assemble SEQ ID NO:13-24 were compared with sequences from the Incyte LIFESEQ database and public domain databases using BLAST and other implementations of the Smith-Waterman algorithm. Sequences from these databases that matched SEQ ID NO:13-24 were assembled into clusters of contiguous and overlapping sequences using assembly algorithms such as Phrap (Table 7). Radiation hybrid and genetic mapping data available from public resources such as the Stanford Human Genome Center (SHGC), Whitehead Institute for Genome Research (WIGR), and Généthon were used to determine if any of the clustered sequences had been previously mapped. Inclusion of a mapped sequence in a cluster resulted in the assignment of all sequences of that cluster, including its particular SEQ ID NO:, to that map location.

Map locations are represented by ranges, or intervals, of human chromosomes. The map position of an interval, in centiMorgans, is measured relative to the terminus of the chromosome's p-arm. (The centiMorgan (cM) is a unit of measurement based on recombination frequencies between chromosomal markers. On average, 1 cM is roughly equivalent to 1 megabase (Mb) of DNA in humans, although this can vary widely due to hot and cold spots of recombination.) The cM distances are based on genetic markers mapped by Généthon which provide boundaries for radiation hybrid markers whose sequences were included in each of the clusters. Human genome maps and other resources available to the public, such as the NCBI "GeneMap'99" World Wide Web site (http://www.ncbi.nlm.nih.gov/genemap/), can be employed to determine if previously identified disease genes map within or in proximity to the intervals indicated above.

VII. Analysis of Polynucleotide Expression

Northern analysis is a laboratory technique used to detect the presence of a transcript of a gene and involves the hybridization of a labeled nucleotide sequence to a membrane on which RNAs from a particular cell type or tissue have been bound. (See, e.g., Sambrook, supra, ch. 7; Ausubel (1995) supra, ch. 4 and 16.)

Analogous computer techniques applying BLAST were used to search for identical or related molecules in cDNA databases such as GenBank or LIFESEQ (Incyte Genomics). This analysis is much faster than multiple membrane-based hybridizations. In addition, the sensitivity of the computer search can be modified to determine whether any particular match is categorized as exact or similar. The basis of the search is the product score, which is defined as:

$$\frac{\text{BLAST Score x Percent Identity}}{\text{5 x minimum \{length(Seq. 1), length(Seq. 2)\}}}$$

The product score takes into account both the degree of similarity between two sequences and the length of the sequence match. The product score is a normalized value between 0 and 100, and is calculated as follows: the BLAST score is multiplied by the percent nucleotide identity and the product is divided by (5 times the length of the shorter of the two sequences). The BLAST score is calculated by assigning a score of +5 for every base that matches in a high-scoring segment pair (HSP), and -4 for every mismatch. Two sequences may share more than one HSP (separated by gaps). If there is more than one HSP, then the pair with the highest BLAST score is used to calculate the product score. The product score represents a balance between fractional overlap and quality in a BLAST alignment. For example, a product score of 100 is produced only for 100% identity over the entire length of the shorter of the two sequences being compared. A product score of 70 is produced either by 100% identity and 70% overlap at one end, or by 88% identity and 100% overlap at the other. A product score of 50 is produced either by 100% identity and 50% overlap at one end, or 79% identity and 100% overlap.

Alternatively, polynucleotide sequences encoding PP are analyzed with respect to the tissue sources from which they were derived. For example, some full length sequences are assembled, at least in part, with overlapping Incyte cDNA sequences (see Example III). Each cDNA sequence is derived from a cDNA library constructed from a human tissue. Each human tissue is classified into one of the following organ/tissue categories: cardiovascular system; connective tissue; digestive system; embryonic structures; endocrine system; exocrine glands; genitalia, female; genitalia, male; germ cells; hemic and immune system; liver; musculoskeletal system; nervous system; pancreas; respiratory system; sense organs; skin; stomatognathic system; unclassified/mixed; or urinary tract. The number of libraries in each category is counted and divided by the total number of libraries across all categories. Similarly, each human tissue is classified into one of the following disease/condition categories: cancer, cell line, developmental, inflammation, neurological, trauma, cardiovascular, pooled, and other, and the number of libraries in each category is counted and divided by the total number of libraries across all categories. The resulting percentages reflect the tissue- and disease-specific expression of cDNA encoding PP. cDNA sequences and cDNA library/tissue information are found in the LIFESEQ GOLD database (Incyte Genomics, Palo Alto CA).

VIII. Extension of PP Encoding Polynucleotides

Full length polynucleotide sequences were also produced by extension of an appropriate fragment of the full length molecule using oligonucleotide primers designed from this fragment. One primer was synthesized to initiate 5' extension of the known fragment, and the other primer was synthesized to initiate 3' extension of the known fragment. The initial primers were designed using OLIGO 4.06 software (National Biosciences), or another appropriate program, to be about 22 to 30 nucleotides in length, to have a GC content of about 50% or more, and to anneal to the target sequence at temperatures of about 68°C to about 72°C. Any stretch of nucleotides which would result in hairpin structures and primer-primer dimerizations was avoided.

Selected human cDNA libraries were used to extend the sequence. If more than one extension was necessary or desired, additional or nested sets of primers were designed.

High fidelity amplification was obtained by PCR using methods well known in the art. PCR was performed in 96-well plates using the PTC-200 thermal cycler (MJ Research, Inc.). The reaction mix contained DNA template, 200 nmol of each primer, reaction buffer containing Mg^{2+}, $(NH_4)_2SO_4$, and 2-mercaptoethanol, Taq DNA polymerase (Amersham Pharmacia Biotech), ELONGASE enzyme (Life Technologies), and Pfu DNA polymerase (Stratagene), with the following parameters for primer pair PCI A and PCI B: Step 1: 94°C, 3 min; Step 2: 94°C, 15 sec; Step 3: 60°C, 1 min; Step 4: 68°C, 2 min; Step 5: Steps 2, 3, and 4 repeated 20 times; Step 6: 68°C, 5 min; Step 7: storage at 4°C. In the alternative, the parameters for primer pair T7 and SK+ were as follows: Step 1: 94°C, 3 min; Step 2: 94°C, 15 sec; Step 3: 57°C, 1 min; Step 4: 68°C, 2 min; Step 5: Steps 2, 3, and 4 repeated 20 times; Step 6: 68°C, 5 min; Step 7: storage at 4°C.

The concentration of DNA in each well was determined by dispensing 100 μl PICOGREEN quantitation reagent (0.25% (v/v) PICOGREEN; Molecular Probes, Eugene OR) dissolved in 1X TE and 0.5 μl of undiluted PCR product into each well of an opaque fluorimeter plate (Corning Costar, Acton MA), allowing the DNA to bind to the reagent. The plate was scanned in a Fluoroskan II (Labsystems Oy, Helsinki, Finland) to measure the fluorescence of the sample and to quantify the concentration of DNA. A 5 μl to 10 μl aliquot of the reaction mixture was analyzed by electrophoresis on a 1 % agarose gel to determine which reactions were successful in extending the sequence.

The extended nucleotides were desalted and concentrated, transferred to 384-well plates, digested with CviJI cholera virus endonuclease (Molecular Biology Research, Madison WI), and sonicated or sheared prior to religation into pUC 18 vector (Amersham Pharmacia Biotech). For shotgun sequencing, the digested nucleotides were separated on low concentration (0.6 to 0.8%) agarose gels, fragments were excised, and agar digested with Agar ACE (Promega). Extended clones were religated using T4 ligase (New England Biolabs, Beverly MA) into pUC 18 vector (Amersham Pharmacia Biotech), treated with Pfu DNA polymerase (Stratagene) to fill-in restriction site overhangs, and transfected into competent E. coli cells. Transformed cells were selected on antibiotic-containing media, and individual colonies were picked and cultured overnight at 37°C in 384-well plates in LB/2x carb liquid media.

The cells were lysed, and DNA was amplified by PCR using Taq DNA polymerase (Amersham Pharmacia Biotech) and Pfu DNA polymerase (Stratagene) with the following parameters: Step 1: 94°C, 3 min; Step 2: 94°C, 15 sec; Step 3: 60°C, 1 min; Step 4: 72°C, 2 min; Step 5: steps 2, 3, and 4 repeated 29 times; Step 6: 72°C, 5 min; Step 7: storage at 4°C. DNA was quantified by PICOGREEN reagent (Molecular Probes) as described above. Samples with low DNA recoveries were reamplified using the same conditions as described above. Samples were diluted with 20% dimethysulfoxide (1:2, v/v), and sequenced using DYENAMIC energy transfer sequencing primers and the DYENAMIC DIRECT kit (Amersham Pharmacia Biotech) or the ABI PRISM BIGDYE Terminator cycle sequencing ready reaction kit (Applied Biosystems).

In like manner, full length polynucleotide sequences are verified using the above procedure or are used to obtain 5' regulatory sequences using the above procedure along with oligonucleotides designed for such extension, and an appropriate genomic library.

IX. Labeling and Use of Individual Hybridization Probes

Hybridization probes derived from SEQ ID NO:13-24 are employed to screen cDNAs, genomic DNAs, or mRNAs. Although the labeling of oligonucleotides, consisting of about 20 base pairs, is specifically described, essentially the same procedure is used with larger nucleotide fragments. Oligonucleotides are designed using state-of-the-art software such as OLIGO 4.06 software (National Biosciences) and labeled by combining 50 pmol of each oligomer, 250 μCi of [γ-^{32}P] adenosine triphosphate (Amersham Pharmacia Biotech), and T4 polynucleotide kinase (DuPont NEN, Boston MA). The labeled oligonucleotides are substantially purified using a SEPHADEX G-25 superfine size exclusion dextran bead column (Amersham Pharmacia Biotech). An aliquot containing 10^7 counts per minute of the labeled probe is used in a typical membrane-based hybridization analysis of human genomic DNA digested with one of the following endonucleases: Ase I, Bgl II, Eco RI, Pst I, Xba I, or Pvu II (DuPont NEN).

The DNA from each digest is fractionated on a 0.7% agarose gel and transferred to nylon membranes (Nytran Plus, Schleicher & Schuell, Durham NH). Hybridization is carried out for 16 hours at 40°C. To remove nonspecific signals, blots are sequentially washed at room temperature under conditions of up to, for example, 0.1 x saline sodium citrate and 0.5% sodium dodecyl sulfate. Hybridization patterns are visualized using autoradiography or an alternative imaging means and compared.

X. Microarrays

The linkage or synthesis of array elements upon a microarray can be achieved utilizing photolithography, piezoelectric printing (ink-jet printing, See, e.g., Baldeschweiler, supra.), mechanical microspotting technologies, and derivatives thereof. The substrate in each of the aforementioned technologies should be uniform and solid with a non-porous surface (Schena (1999), supra). Suggested substrates include silicon, silica, glass slides, glass chips, and silicon wafers. Alternatively, a procedure analogous to a dot or slot blot may also be used to arrange and link elements to the surface of a substrate using thermal, UV, chemical, or mechanical bonding procedures. A typical array may be produced using available methods and machines well known to those of ordinary skill in the art and may contain any appropriate number of elements. (See, e.g., Schena, M. et al. (1995) Science 270:467-470; Shalon, D. et al. (1996) Genome Res. 6:639-645; Marshall, A. and J. Hodgson (1998) Nat. Biotechnol. 16:27-31.)

Full length cDNAs, Expressed Sequence Tags (ESTs), or fragments or oligomers thereof may comprise the elements of the microarray. Fragments or oligomers suitable for hybridization can be selected using software well known in the art such as LASERGENE software (DNASTAR). The array elements are hybridized with polynucleotides in a biological sample. The polynucleotides in the biological sample are conjugated to a fluorescent label or other molecular tag for ease of detection. After hybridization, nonhybridized nucleotides from the biological sample are removed, and a fluorescence scanner is used to detect hybridization at each array element. Alternatively, laser desorbtion and mass spectrometry may be used for detection of hybridization. The degree of complementarity and the relative abundance of each polynucleotide which hybridizes to an element on the microarray may be assessed. In one embodiment, microarray preparation and usage is described in detail below.

Tissue or Cell Sample Preparation

Total RNA is isolated from tissue samples using the guanidinium thiocyanate method and poly(A)$^+$ RNA is purified using the oligo-(dT) cellulose method. Each poly(A)$^+$ RNA sample is reverse transcribed using MMLV reverse-transcriptase, 0.05 pg/μl oligo-(dT) primer (21mer), 1X first strand buffer, 0.03 units/μl RNase inhibitor, 500 μM dATP, 500 μM dGTP, 500 μM dTTP, 40 μM dCTP, 40 μM dCTP-Cy3 (BDS) or dCTP-Cy5 (Amersham Pharmacia Biotech). The reverse transcription reaction is performed in a 25 ml volume containing 200 ng poly(A)$^+$ RNA with GEMBRIGHT kits (Incyte). Specific control poly(A)$^+$ RNAs are synthesized by in vitro transcription from non-coding yeast genomic DNA. After incubation at 37°C for 2 hr, each reaction sample (one with Cy3 and another with Cy5 labeling) is treated with 2.5 ml of 0.5M sodium hydroxide and incubated for 20 minutes at 85° C to the stop the reaction and degrade the RNA. Samples are purified using two successive CHROMA SPIN 30 gel filtration spin columns (CLONTECH Laboratories, Inc. (CLONTECH), Palo Alto CA) and after combining, both reaction samples are ethanol precipitated using 1 ml of glycogen (1 mg/ml), 60 ml sodium acetate, and 300 ml of 100% ethanol. The sample is then dried to completion using a SpeedVAC (Savant Instruments Inc., Holbrook NY) and resuspended in 14 μl 5X SSC/0.2% SDS.

Microarray Preparation

Sequences of the present invention are used to generate array elements. Each array element is amplified from bacterial cells containing vectors with cloned cDNA inserts. PCR amplification uses primers complementary to the vector sequences flanking the cDNA insert. Array elements are amplified in thirty cycles of PCR from an initial quantity of 1-2 ng to a final quantity greater than 5 μg. Amplified array elements are then purified using SEPHACRYL-400 (Amersham Pharmacia Biotech).

Purified array elements are immobilized on polymer-coated glass slides. Glass microscope slides (Corning) are cleaned by ultrasound in 0.1% SDS and acetone, with extensive distilled water washes between and after treatments. Glass slides are etched in 4% hydrofluoric acid (VWR Scientific Products Corporation (VWR), West Chester PA), washed extensively in distilled water, and coated with 0.05% aminopropyl silane (Sigma) in 95% ethanol. Coated slides are cured in a 110°C oven.

Array elements are applied to the coated glass substrate using a procedure described in US Patent No. 5,807,522, incorporated herein by reference. 1 μl of the array element DNA, at an average concentration of 100 ng/μl, is loaded into the open capillary printing element by a high-speed robotic apparatus. The apparatus then deposits about 5 nl of array element sample per slide.

Microarrays are UV-crosslinked using a STRATALINKER UV-crosslinker (Stratagene). Microarrays are washed at room temperature once in 0.2% SDS and three times in distilled water. Non-specific binding sites are blocked by incubation of microarrays in 0.2% casein in phosphate buffered saline (PBS) (Tropix, Inc., Bedford MA) for 30 minutes at 60° C followed by washes in 0.2% SDS and distilled water as before.

Hybridization

Hybridization reactions contain 9 μl of sample mixture consisting of 0.2 μg each of Cy3 and Cy5 labeled cDNA synthesis products in 5X SSC, 0.2% SDS hybridization buffer. The sample mixture is heated to 65° C for 5 minutes and is aliquoted onto the microarray surface and covered with an 1.8 cm^2 coverslip. The arrays are transferred to a waterproof chamber having a cavity just slightly larger than a microscope slide. The chamber is kept at 100% humidity internally by the addition of 140 µl of 5X SSC in a corner of the chamber. The chamber containing the arrays is incubated for about 6.5 hours at 60° C. The arrays are washed for 10 min at 45° C in a first wash buffer (1X SSC, 0.1% SDS), three times for 10 minutes each at 45° C in a second wash buffer (0.1X SSC), and dried.

Detection

Reporter-labeled hybridization complexes are detected with a microscope equipped with an Innova 70 mixed gas 10 W laser (Coherent, Inc., Santa Clara CA) capable of generating spectral lines at 488 nm for excitation of Cy3 and at 632 nm for excitation of Cy5. The excitation laser light is focused on the array using a 20X microscope objective (Nikon, Inc., Melville NY). The slide containing the array is placed on a computer-controlled X-Y stage on the microscope and raster-scanned past the objective. The 1.8 cm x 1.8 cm array used in the present example is scanned with a resolution of 20 micrometers.

In two separate scans, a mixed gas multiline laser excites the two fluorophores sequentially. Emitted light is split, based on wavelength, into two photomultiplier tube detectors (PMT R1477, Hamamatsu Photonics Systems, Bridgewater NJ) corresponding to the two fluorophores. Appropriate filters positioned between the array and the photomultiplier tubes are used to filter the signals. The emission maxima of the fluorophores used are 565 nm for Cy3 and 650 nm for Cy5. Each array is typically scanned twice, one scan per fluorophore using the appropriate filters at the laser source, although the apparatus is capable of recording the spectra from both fluorophores simultaneously.

The sensitivity of the scans is typically calibrated using the signal intensity generated by a cDNA control species added to the sample mixture at a known concentration. A specific location on the array contains a complementary DNA sequence, allowing the intensity of the signal at that location to be correlated with a weight ratio of hybridizing species of 1:100,000. When two samples from different sources (e.g., representing test and control cells), each labeled with a different fluorophore, are hybridized to a single array for the purpose of identifying genes that are differentially expressed, the calibration is done by labeling samples of the calibrating cDNA with the two fluorophores and adding identical amounts of each to the hybridization mixture.

The output of the photomultiplier tube is digitized using a 12-bit RTI-835H analog-to-digital (A/D) conversion board (Analog Devices, Inc., Norwood MA) installed in an IBM-compatible PC computer. The digitized data are displayed as an image where the signal intensity is mapped using a linear 20-color transformation to a pseudocolor scale ranging from blue (low signal) to red (high signal). The data is also analyzed quantitatively. Where two different fluorophores are excited and measured simultaneously, the data are first corrected for optical crosstalk (due to overlapping emission spectra) between the fluorophores using each fluorophore's emission spectrum.

A grid is superimposed over the fluorescence signal image such that the signal from each spot is centered in each element of the grid. The fluorescence signal within each element is then integrated to obtain a numerical value corresponding to the average intensity of the signal. The software used for signal analysis is the GEMTOOLS gene expression analysis program (Incyte).

XI. Complementary Polynucleotides

Sequences complementary to the PP-encoding sequences, or any parts thereof, are used to detect, decrease, or inhibit expression of naturally occurring PP. Although use of oligonucleotides comprising from about 15 to 30 base pairs is described, essentially the same procedure is used with smaller or with larger sequence fragments. Appropriate oligonucleotides are designed using OLIGO 4.06 software (National Biosciences) and the coding sequence of PP. To inhibit transcription, a complementary oligonucleotide is designed from the most unique 5' sequence and used to prevent promoter binding to the coding sequence. To inhibit translation, a complementary oligonucleotide is designed to prevent ribosomal binding to the PP-encoding transcript.

XII. Expression of PP

Expression and purification of PP is achieved using bacterial or virus-based expression systems. For expression of PP in bacteria, cDNA is subcloned into an appropriate vector containing an antibiotic resistance gene and an inducible promoter that directs high levels of cDNA transcription. Examples of such promoters include, but are not limited to, the *trp-lac* (*tac*) hybrid promoter and the T5 or T7 bacteriophage promoter in conjunction with the *lac* operator regulatory element. Recombinant vectors are transformed into suitable bacterial hosts, e.g., BL21(DE3). Antibiotic resistant bacteria express PP upon induction with isopropyl beta-D-thiogalactopyranoside (IPTG). Expression of PP in eukaryotic cells is achieved by infecting insect or mammalian cell lines with recombinant Autographica californica nuclear polyhedrosis virus (AcMNPV), commonly known as baculovirus. The nonessential polyhedrin gene of baculovirus is replaced with cDNA encoding PP by either homologous recombination or bacterial-mediated transposition involving transfer plasmid intermediates. Viral infectivity is maintained and the strong polyhedrin promoter drives high levels of cDNA transcription. Recombinant baculovirus is used to infect Spodoptera frugiperda (Sf9) insect cells in most cases, or human hepatocytes, in some cases. Infection of the latter requires additional genetic modifications to baculovirus. (See Engelhard, E.K. et al. (1994) Proc. Natl. Acad. Sci. USA 91:3224-3227; Sandig, V. et al. (1996) Hum. Gene Ther. 7:1937-1945.)

In most expression systems, PP is synthesized as a fusion protein with, e.g., glutathione S-transferase (GST) or a peptide epitope tag, such as FLAG or 6-His, permitting rapid, single-step, affinity-based purification of recombinant fusion protein from crude cell lysates. GST, a 26-kilodalton enzyme from Schistosoma japonicum, enables the purification of fusion proteins on immobilized glutathione under conditions that maintain protein activity and antigenicity (Amersham

Pharmacia Biotech). Following purification, the GST moiety can be proteolytically cleaved from PP at specifically engineered sites. FLAG, an 8-amino acid peptide, enables immunoaffinity purification using commercially available monoclonal and polyclonal anti-FLAG antibodies (Eastman Kodak). 6-His, a stretch of six consecutive histidine residues, enables purification on metal-chelate resins (QIAGEN). Methods for protein expression and purification are discussed in Ausubel (1995, supra, ch. 10 and 16). Purified PP obtained by these methods can be used directly in the assays shown in Examples XVI, XVII, XVIII, and XIX where applicable.

XIII. Functional Assays

PP function is assessed by expressing the sequences encoding PP at physiologically elevated levels in mammalian cell culture systems. cDNA is subcloned into a mammalian expression vector containing a strong promoter that drives high levels of cDNA expression. Vectors of choice include PCMV SPORT (Life Technologies) and PCR3.1 (Invitrogen, Carlsbad CA), both of which contain the cytomegalovirus promoter. 5-10 μg of recombinant vector are transiently transfected into a human cell line, for example, an endothelial or hematopoietic cell line, using either liposome formulations or electroporation. 1-2 μg of an additional plasmid containing sequences encoding a marker protein are co-transfected. Expression of a marker protein provides a means to distinguish transfected cells from nontransfected cells and is a reliable predictor of cDNA expression from the recombinant vector. Marker proteins of choice include, e.g., Green Fluorescent Protein (GFP; Clontech), CD64, or a CD64-GFP fusion protein. Flow cytometry (FCM), an automated, laser optics-based technique, is used to identify transfected cells expressing GFP or CD64-GFP and to evaluate the apoptotic state of the cells and other cellular properties. FCM detects and quantifies the uptake of fluorescent molecules that diagnose events preceding or coincident with cell death. These events include changes in nuclear DNA content as measured by staining of DNA with propidium iodide; changes in cell size and granularity as measured by forward light scatter and 90 degree side light scatter; down-regulation of DNA synthesis as measured by decrease in bromodeoxyuridine uptake; alterations in expression of cell surface and intracellular proteins as measured by reactivity with specific antibodies; and alterations in plasma membrane composition as measured by the binding of fluorescein-conjugated Annexin V protein to the cell surface. Methods in flow cytometry are discussed in Ormerod, M.G. (1994) Flow Cytometry, Oxford, New York NY.

The influence of PP on gene expression can be assessed using highly purified populations of cells transfected with sequences encoding PP and either CD64 or CD64-GFP. CD64 and CD64-GFP are expressed on the surface of transfected cells and bind to conserved regions of human immunoglobulin G (IgG). Transfected cells are efficiently separated from nontransfected cells using magnetic beads coated with either human IgG or antibody against CD64 (DYNAL, Lake Success NY). mRNA can be purified from the cells using methods well known by those of skill in the art.

Expression of mRNA encoding PP and other genes of interest can be analyzed by northern analysis or microarray techniques.

XIV. Production of PP Specific Antibodies

PP substantially purified using polyacrylamide gel electrophoresis (PAGE; see, e.g., Harrington, M.G. (1990) Methods Enzymol. 182:488-495), or other purification techniques, is used to immunize rabbits and to produce antibodies using standard protocols.

Alternatively, the PP amino acid sequence is analyzed using LASERGENE software (DNASTAR) to determine regions of high immunogenicity, and a corresponding oligopeptide is synthesized and used to raise antibodies by means known to those of skill in the art. Methods for selection of appropriate epitopes, such as those near the C-terminus or in hydrophilic regions are well described in the art. (See, e.g., Ausubel, 1995, supra, ch. 11.)

Typically, oligopeptides of about 15 residues in length are synthesized using an ABI 431A peptide synthesizer (Applied Biosystems) using FMOC chemistry and coupled to KLH (Sigma-Aldrich, St. Louis MO) by reaction with N-maleimidobenzoyl-N-hydroxysuccinimide ester (MBS) to increase immunogenicity. (See, e.g., Ausubel, 1995, supra.) Rabbits are immunized with the oligopeptide-KLH complex in complete Freund's adjuvant. Resulting antisera are tested for antipeptide and anti-PP activity by, for example, binding the peptide or PP to a substrate, blocking with 1% BSA, reacting with rabbit antisera, washing, and reacting with radio-iodinated goat anti-rabbit IgG.

XV. Purification of Naturally Occurring PP Using Specific Antibodies

Naturally occurring or recombinant PP is substantially purified by immunoaffinity chromatography using antibodies specific for PP. An immunoaffinity column is constructed by covalently coupling anti-PP antibody to an activated chromatographic resin, such as CNBr-activated SEPHAROSE (Amersham Pharmacia Biotech). After the coupling, the resin is blocked and washed according to the manufacturer's instructions.

Media containing PP are passed over the immunoaffinity column, and the column is washed under conditions that allow the preferential absorbance of PP (e.g., high ionic strength buffers in the presence of detergent). The column is eluted under conditions that disrupt antibody/PP binding (e.g., a buffer of pH 2 to pH 3, or a high concentration of a chaotrope, such as urea or thiocyanate ion), and PP is collected.

XVI. Identification of Molecules Which Interact with PP

PP, or biologically active fragments thereof, are labeled with ^{125}I Bolton-Hunter reagent. (See, e.g., Bolton A.E. and W.M. Hunter (1973) Biochem. J. 133:529-539.) Candidate molecules previously arrayed in the wells of a multi-well plate are incubated with the labeled PP, washed, and any wells with labeled PP complex are assayed. Data obtained using different concentrations of PP are used to calculate values for the number, affinity, and association of PP with the candidate molecules.

Alternatively, molecules interacting with PP are analyzed using the yeast two-hybrid system as described in Fields, S. and O. Song (1989) Nature 340:245-246, or using commercially available kits based on the two-hybrid system, such as the MATCHMAKER system (Clontech).

PP may also be used in the PATHCALLING process (CuraGen Corp., New Haven CT) which employs the yeast two-hybrid system in a high-throughput manner to determine all interactions between the proteins encoded by two large libraries of genes (Nandabalan, K. et al. (2000) U.S. Patent No. 6,057,101).

XVII. Demonstration of PP Activity

PP activity is measured by the hydrolysis of para-nitrophenyl phosphate (PNPP). PP is incubated together with PNPP in HEPES buffer pH 7.5, in the presence of 0.1% β-mercaptoethanol at 37°C for 60 min. The reaction is stopped by the addition of 6 ml of 10 N NaOH (Diamond, R.H. et al. (1994) Mol. Cell. Biol. 14:3752-62). Alternatively, acid phosphatase activity of PP is demonstrated by incubating PP-containing extract with 100 μl of 10 mM PNPP in 0.1 M sodium citrate, pH 4.5, and 50 μl of 40 mM NaCl at 37°C for 20 min. The reaction is stopped by the addition of 0.5 ml of 0.4 M glycine/NaOH, pH 10.4 (Saftig, P. et al. (1997) J. Biol. Chem. 272:18628-18635). The increase in light absorbance at 410 nm resulting from the hydrolysis of PNPP is measured using a spectrophotometer. The increase in light absorbance is proportional to the activity of PP in the assay.

In the alternative, PP activity is determined by measuring the amount of phosphate removed from a phosphorylated protein substrate. Reactions are performed with 2 or 4 nM enzyme in a final volume of 30 μl containing 60 mM Tris, pH 7.6, 1 mM EDTA, 1 mM EGTA, 0.1% β-mercaptoethanol and 10 μM substrate, ^{32}P-labeled on serine/threonine or tyrosine, as appropriate. Reactions are initiated with substrate and incubated at 30° C for 10-15 min. Reactions are quenched with 450 μl of 4% (w/v) activated charcoal in 0.6 M HCl, 90 mM $Na_4P_2O_7$, and 2 mM NaH_2PO_4, then centrifuged at 12,000 × *g* for 5 min. Acid-soluble ^{32}Pi is quantified by liquid scintillation counting (Sinclair, C. et al. (1999) J. Biol. Chem. 274:23666-23672).

XVIII. Identification of PP Inhibitors

Compounds to be tested are arrayed in the wells of a 384-well plate in varying concentrations along with an appropriate buffer and substrate, as described in the assays in Example XVII. PP activity is measured for each well and the ability of each compound to inhibit PP activity can be determined, as well as the dose-response kinetics. This assay could also be used to identify molecules which enhance PP activity.

XIX. Identification of PP Substrates

A PP "substrate-trapping" assay takes advantage of the increased substrate affinity that may be conferred by certain mutations in the PTP signature sequence. PP bearing these mutations form a stable complex with their substrate; this complex may be isolated biochemically. Site-directed mutagenesis of invariant residues in the PTP signature sequence in a clone encoding the catalytic domain of PP is performed using a method standard in the art or a commercial kit, such as the MUTA-GENE kit from BIO-RAD. For expression of PP mutants in Escherichia coli, DNA fragments containing the mutation are exchanged with the corresponding wild-type sequence in an expression vector bearing the sequence encoding PP or a glutathione S-transferase (GST)-PP fusion protein. PP mutants are expressed in E. coli and purified by chromatography.

The expression vector is transfected into COS1 or 293 cells via calcium phosphate-mediated transfection with 20 μg of CsCl-purified DNA per 10-cm dish of cells or 8 μg per 6-cm dish. Forty-eight hours after transfection, cells are stimulated with 100 ng/ml epidermal growth factor to increase tyrosine phosphorylation in cells, as the tyrosine kinase EGFR is abundant in COS cells. Cells are lysed in 50 mM Tris·HCl, pH 7.5/5 mM EDTA/150 mM NaCl/1% Triton X-100/5 mM iodoacetic acid/10 mM sodium phosphate/10 mM NaF/5 μg/ml leupeptin/5 μg/ml aprotinin/1 mM benzamidine (1 ml per 10-cm dish, 0.5 ml per 6-cm dish). PP is immunoprecipitated from lysates with an appropriate antibody. GST-PP fusion proteins are precipitated with glutathione-Sepharose, 4 μg of mAb or 10 μl of beads respectively per mg of cell lysate. Complexes can be visualized by PAGE or further purified to identify substrate molecules (Flint, A.J. et al. (1997) Proc. Natl. Acad. Sci. USA 94:1680-1685).

Various modifications and variations of the described methods and systems of the invention will be apparent to those skilled in the art without departing from the scope and spirit of the invention. Although the invention has been described in connection with certain embodiments, it should be understood that the invention as claimed should not be unduly limited to such specific embodiments. Indeed, various modifications of the described modes for carrying out the invention which are obvious to those skilled in molecular biology or related fields are intended to be within the scope of the following claims.

Table 1

Incyte Project ID	Polypeptide SEQ ID NO:	Incyte Polypeptide ID	Polynucleotide SEQ ID NO:	Incyte Polynucleotide ID
3272350	1	3272350CD1	13	3272350CB1
7481507	2	7481507CD1	14	7481507CB1
2285140	3	2285140CD1	15	2285140CB1
7197873	4	7197873CD1	16	7197873CB1
6282188	5	6282188CD1	17	6282188CB1
2182961	6	2182961CD1	18	2182961CB1
5119906	7	5119906CD1	19	5119906CB1
4022502	8	4022502CD1	20	4022502CB1
4084356	9	4084356CD1	21	4084356CB1
1740204	10	1740204CD1	22	1740204CB1
7483804	11	7483804CD1	23	7483804CB1
7483934	12	7483934CD1	24	7483934CB1

Table 2

Polypeptide SEQ ID NO:	Incyte Polypeptide ID	GenBank ID NO:	Probability score	GenBank Homolog
1	3272350CD1	g1418932	1.60E-14	[Homo sapiens] human phosphotyrosine phosphatase kappa Fuchs, M. et al. (1996) J. Biol. Chem. 271(28):16712-16719
2	7481507CD1	g13360272	4.00E-95	[Escherichia coli O157:H7] serine/threonine protein phosphatase Makino, K. et al. (1999) Genes Genet. Syst. 74(5):227-239
3	2285140CD1	g3874135	7.60E-55	[Caenorhabditis elegans] similar to acid phosphatase
4	7197873CD1	g452194	2.60E-169	[Homo sapiens] protein tyrosine phosphatase (PTP-BAS, type 3) Maekawa, K., et al. (1994) FEBS Lett. 337:200-206
5	6282188CD1	g9759130	1.80E-06	[Arabidopsis thaliana] contains similarity to tyrosine gene_id:MZK4.21
6	2182961CD1	g3876155	3.50E-84	[Caenorhabditis elegans] Similar to Aspergillus acid phosphatase
7	5119906CD1	g6714641	7.30E-101	[Drosophila melanogaster] MAP kinase phosphatase
8	4022502CD1	g12746390	1.00E-46	[Rattus norvegicus] sphingosine-1-phosphate phosphohydrolase
9	4084356CD1	g3063745	3.50E-77	[Bos taurus] protein Phosphatase 2C beta Klumpp, S. et al. (1998) J. Neurosci. Res. 51:328-338
10	1740204CD1	g619215	7.40E-244	[Oryctolagus cuniculus] protein phosphatase 2A1 B gamma subunit Zolnierowicz, S. et al. (1994) Biochem. 33:11858-11867
11	7483804CD1	g957217	2.80E-292	[Homo sapiens] striatum-enriched phosphatase Li, X. (1995) Genomics 28:442-449
12	7483934CD1	g4104822	0	[Homo sapiens] synaptojanin 2B Nemoto, Y and De Camilli, P. (1999) EMBO J. 18(11):2991-3006

Table 3

SEQ ID NO:	Incyte Polypeptide ID	Amino Acid Residues	Potential Phosphorylation Sites	Potential Glycosylation Sites	Signature Sequences, Domains and Motifs	Analytical Methods and Databases
1	3272350CD1	435	S111 S129 S159 S24 S308 T143 T288 T390 T401	N135 N227 N306	PROTEIN-TYROSINE-PHOSPHATASE, RECEPTOR TYPE MU DM07136|P35822|1-187:C233-V389 MAM DM01344|P28824|595-796:S229-D387	BLAST_DOMO
					PRECURSOR GLYCOPROTEIN SIGNAL TRANSMEMBRANE HYDROLASE PROTEIN REPEAT RECEPTOR PHOSPHATASE NEUROPILIN PD001482:D230-C396	BLAST_PRODOM
					MAM domain proteins. BL00740A:C241-W253, BL00740B:L381-T401	BLIMPS_BLOCKS
					MAM domain signature PR00020A:K239-N257, PR00020C:Y312-K323, PR00020D:V360-G374, PR00020E:G379-K392	BLIMPS_PRINT
					MAM domain. MAM:C233-R398 Immunoglobulin domain. ig:G33-V97; C241-S315	HMMER_PFAM
					Spscan signal_cleavage:M1-P47	SPSCAN
2	7481507CD1	233	S57 T4	N221	SERINE/THREONINE PROTEIN PHOSPHATASE HYDROLASE IRON MANGANESE PD152367:Q89-Q228	BLAST_PRODOM
					Ser_Thr_Phosphatase V83-E88	MOTIFS
					Serine/threonine specific protein phosphatases signature ser_thr_phosphatase.prf:D63-G108	PROFILESCAN
3	2285140CD1	315	S12 S168 S20 S21 S215 S229 S244 S251 S252 S257 S266 S267 S278 S283 S285 S301 T104 T236 T27 Y87		PROTEIN PHYB1 PUTATIVE ACID PHOSPHATASE F26C11.1 HYDROLASE PD146082: D57-L315	BLAST_PRODOM

Table 3 (cont.)

SEQ ID NO:	Incyte Polypeptide ID	Amino Acid Residues	Potential Phosphorylation Sites	Potential Glycosylation Sites	Signature Sequences, Domains and Motifs	Analytical Methods and Databases
4	7197873CD1	1278	S1013 S1017 S1040 S1177 S1197 S122 S1254 S157 S20 S204 S212 S249 S255 S266 S278 S471 S523 S589 S649 S651 S68 S733 S85 S850 S908 S913 T108 T1162 T119 T1242 T155 T185 T301 T562 T573 T579 T584 T592 T670 T694 T746 T788 T981 Y430 Y461	N1015 N173 N41 N548 N842 N938	Band 4.1 family domain signatures:A494-E545	PROFILESCAN
					BAND 4 DM00609|A54971|562-990: T301-Q614	BLAST_DOMO
					BAND 4 DM00609|S51005|13-453: T301-F606	BLAST_DOMO
					BAND 4 DM00609|JC4155|11-447: K305-Q609	BLAST_DOMO
					GLGF DOMAIN DM00224|A54971|1358-1454: S908-E1001	BLAST_DOMO
					PROTEIN CYTOSKELETON STRUCTURAL PHOSPHATASE HYDROLASE PROTEIN TYROSINE PHOSPHORYLATION MOESIN TYROSINE BAND PD000961: L313-V516	BLAST_PRODOM
					PHOSPHATASE TYROSINE PROTEIN TYPE PTP BAS HYDROLASE PROTEIN TYROSINE PHOSPHATASE PHOSPHOTYROSINE PTPASE 1E PD008840: V6-S85	BLAST_PRODOM
					PHOSPHATASE TYROSINE PROTEIN TYPE PTP BAS HYDROLASE PROTEIN TYROSINE PHOSPHATASE PHOSPHOTYROSINE PTPASE 1E PD150192: H519-Q614	BLAST_PRODOM
					Band 4.1 family domain BL00660D: F557-G580, F585-N607, G320-C372, R413-P452, D499-I542	BLIMPS_BLOCKS
					BAND 4.1 PROTEIN FAMILY PR00935: A344-Y356, L418-C431, C431-Y451, D499-G515	BLIMPS_PRINTS
					signal peptide: M1-A18	HMMER
					FERM domain (Band 4.1 family) Band_41: T392-H519	HMMER_PFAM
					PDZ domain (Also known as DHR or GLGF). PDZ: R744-S829, F919-R1003, E1048-P1135	HMMER_PFAM
5	6282188CD1	218	S134 S152 S204 S31 T212 T99	N123 N20	Tyrosine specific protein phosphatases active site: V76-T161	PROFILESCAN
					signal peptide: M1-A25	HMMER, SPSCAN
					Tyrosine specific protein phosphatase signature: M128-M140	MOTIFS

Table 3 (cont.)

SEQ ID NO:	Incyte Polypeptide ID	Amino Acid Residues	Potential Phosphorylation Sites	Potential Glycosylation Sites	Signature Sequences, Domains and Motifs	Analytical Methods and Databases
6	2182961CD1	420	S100 S115 S274 T139 T282 T38 T79 Y379	N193 N332 N386	Purple acid phosphatase PA_phosphatase: N187-P366	HMMER_PFAM
					ACID PHOSPHATASE PURPLE HYDROLASE IRONIII ZINCII PD006329: M182-D400	BLAST_PRODOM
					PHOSPHATASE II; PURPLE; IRON; DM08310 P80366|75-291: Y74-S274 S51078|1-211: Y74-M266	BLAST_DOMO
					PHOSPHATASE; ACID; DM08309|JC2545|292-446: S178-F230	BLAST_DOMO
7	5119906CD1	986	S169 S21 S262 S270 S378 S452 S458 S513 S559 S582 S585 S657 S676 S69 S697 S715 S734 S786 S791 S831 S873 S880 S925 S931 S969 T145 T170 T183 T188 T209 T527 T543 T684 T763 T937 T945 T958 T965 Y591	N207 N260 N277 N322 N557 N732 N868	Signal_cleavage: M1-C37	SPSCAN
					signal peptide: M39-A62	HMMER
					Inhibin beta C chain signature PR00672 K108-K124	BLIMPS_PRINTS
					Dual specificity phosphatase, catalytic domain DSPc: K245-I383	HMMER_PFAM
					Lymphocyte-specific protein PR01083: E528-Q547	BLIMPS_PRINTS
					Tyrosine specific protein phosphatases signature BL00383: V328-A338	BLIMPS_BLOCKS
					VH1-TYPE DUAL SPECIFICITY PHOSPHATASE DM03823|P28562|169-314: P246-E381 DM08829|P38590|138-376: M243-L384	BLAST_DOMO
8	4022502CD1	399	S118 S180 S7 T214 T237 T39 T273 Y373	N344	Transmembrane domain: I113-Y130, F189-Y209, I280-L299, V322-V342	HMMER
					Magnesium independent phosphatidate phosphatase (PAP2) superfamily: S93-C241	HMMER_PFAM
					Intergenic Region Transmembrane Protein RPS21BMRS3 MRS4DYN1 PD042353: F90-E368	BLAST_PRODOM

Table 3 (cont.)

SEQ ID NO:	Incyte Polypeptide ID	Amino Acid Residues	Potential Phosphorylation Sites	Potential Glycosylation Sites	Signature Sequences, Domains and Motifs	Analytical Methods and Databases
9	4084356CD1	387	S103 S128 S153 S381 S88 T174 T194 Y195	N275	signal_cleavage: M1-A67	SPSCAN
					Protein phosphatase 2C: L22-M276	HMMER_PFAM
					Protein phosphatase 2C proteins BL01032: S272-V281, Q30-H40, L55-G64, G92-R109, G118-V127, H136-I175, R179-D192, D223-D235	BLIMPS_BLOCKS
					PROTEIN PHOSPHATASE 2C MAGNESIUM HYDROLASE MANGANESE MULTIGENE FAMILY PP2C ISOFORM PD001101:E91-R296, L22-T117	BLAST_PRODOM
					PROTEIN PHOSPHATASE 2C DM00377|P36993|1-304:H13-A293 DM00377|S39781|1-304:H13-A293 DM00377|I49016|1-304:H13-A293 DM00377|P35815|1-304:H13-A293	BLAST_DOMO
					ATP/GTP-binding site motif A (P-loop): G367-S374	MOTIFS
10	1740204CD1	447	S109 S163 S183 S190 S242 S246 S272 S28 S283 S292 S331 S381 S63 T114 T121 T204 T22 T226 T303 T369 T412 T7	N11 N273 N33 N347	Protein phosphatase 2A regulatory subunit BL01024: C185-L221, T222-F265, E266-I316, E317-G348, K389-K441, T22-D68, L86-R126, T146-D184	BLIMPS_BLOCKS
					Protein phosphatase 2A regulatory subunit PR00600: E31-F51, E66-K94, I95-R123, H172-W199, H200-A227, S228-A256, L257-V284, S285-E312, A313-I338, F339-F365, L409-F438	BLIMPS_PRINTS
					Protein phosphatase 2A regulatory subunit Pr55_1: E79-N93	MOTIFS
					PROTEIN PHOSPHATASE 2A REGULATORY SUBUNIT DM02681|A55836|1-447: M1-M446 DM02681|P36872|60-498: V21-K441 DM02681|Q00362|1-525: V21-F365, D407-F438 DM02681|S55889|13-513: V145-F438, V21-R123	BLAST_DOMO
					SUBUNIT PP2A PHOSPHATASE REGULATORY PROTEIN B ISOFORM MULTIGENE FAMILY PD004712: N131-R385, D17-Y130, N347-F438 PD004812: D407-F438	BLAST_PRODOM

Table 3 (cont.)

SEQ ID NO:	Incyte Polypeptide ID	Amino Acid Residues	Potential Phosphorylation Sites	Potential Glycosylation Sites	Signature Sequences, Domains and Motifs	Analytical Methods and Databases
					WD domain, G-beta repeat WD40: R16-Q52, Y82-K119, R165-H200, N273-D308, S331-D366, S405-Q439	HMMER_PFAM
11	7483804CD1	572	S184 S245 S268 S273 S28 S448 S451 S465 S52 S560 S8 T168 T192 T233 T342 T386 T406 T83	N115 N246	Protein-tyrosine phosphatase active site Y_phosphatase: L322-L561	HMMER_PFAM
					Protein-tyrosine phosphatase active site Tyr_Phosphatase: V501-F513	MOTIFS
					Tyrosine protein phosphatase active site tyr_phosphatase.prf: L478-R539	PROFILESCAN
					Tyrosine specific protein phosphatase BL00383: R539-F554, K325-V339, S351-I359, D389-T399, H460-P472, V501-G511	BLIMPS_BLOCKS
					Protein tyrosine phosphatase PR00700: S352-I359, Y376-Q396, R456-D473, P498-T516, V529-G544, M545-V555	BLIMPS_PRINTS
					PROTEIN-TYROSINE-PHOSPHATASE DM00089 | P35234 | 89-362: L285-L566 DM00089 | P54830 | 261-534: L285-L566 DM00089 | A55574 | 377-649: L285-L566 DM00089 | A55769 | 133-405: L285-L566	BLAST_DOMO
					PROTEINTYROSINE PHOSPHATASE STRIATUMENRICHED NEURAL SPECIFIC HYDROLASE ALTERNATIVE SPLICE PD099306: M25-W196	BLAST_PRODOM
					PHOSPHATASE PROTEINTYROSINE SIGNAL PRECURSOR TRANSMEMBRANE GLYCOPROTEIN RECEPTOR PD000167: K325-G527 PD000155: R456-Y562	BLAST_PRODOM
					PHOSPHATASE PROTEINTYROSINE SIGNAL PRECURSOR LCPTP HEMATOPOIETIC HEPTP STRIATUM ENRICHED PD005701: K235-G321	BLAST_PRODOM
					transmembrane _domain: L146-L166, I499-L523	HMMER

Table 3 (cont.)

SEQ ID NO:	Incyte Polypeptide ID	Amino Acid Residues	Potential Phosphorylation Sites	Potential Glycosylation Sites	Signature Sequences, Domains and Motifs	Analytical Methods and Databases
12	7483934CD1	1510	S1024 S1046 S1136 S1221 S1265 S1316 S1353 S1457 S1493 S152 S190 S211 S388 S429 S440 S585 S590 S632 S776 S806 S839 S961 S998 T1018 T1074 T1081 T114 T1155 T1256 T1260 T1284 T1404 T1491 T157 T160 T204 T382 T395 T41 T524 T532 T616 T667 T718 T748 T793 T815 T830 T914 T952 Y111	N1376 N1440 N612	Inositol polyphosphate phosphatase family, catalytic domain: K542-D884	HMMER_PFAM
					YOR109W; MEMBRANE; DM02715|P50942|65-597: Q122-W551	BLAST_DOMO
					SYNDROME; YOR109W; OCULOCEREBRORENAL; MEMBRANE; DM02714|Q01968|323-658:D588-W820	BLAST_DOMO
					SYNDROME; YOR109W; OCULOCEREBRORENAL; MEMBRANE; DM02714|P50942|599-979:N552-D838	BLAST_DOMO
					SYNDROME; YOR109W; OCULOCEREBRORENAL; MEMBRANE; DM02714|S61667|574-958:D593-W820	BLAST_DOMO
					KIAA0348 PD142428: P1266-T1510	BLAST_PRODOM
					KIAA0348 SYNAPTOJANIN ISOFORM ALPHA PD155999: F1040-S1265	BLAST_PRODOM
					PROTEIN INOSITOL HYDROLASE 5-PHOSPHATASE SYNAPTOJANIN POLYPHOSPHATE PHOSPHATASE TYPE I POLYPHOSPHATE 5-PHOSPHATASE PD002029: D587-D884	BLAST_PRODOM
					SYNAPTOJANIN ENDOCYTOSIS KIAA0348 II ISOFORM ALPHA DELTASACSYNAPTOJANIN1 PD011649:R888-P1128	BLAST_PRODOM
					Inositol polyphosphate phosphatase family, catalytic domain PF00783: F736-I745, R810-L819	BLIMPS_PFAM

Table 4

Polynucleotide SEQ ID NO:	Incyte Polynucleotide ID	Sequence Length	Selected Fragment(s)	Sequence Fragments	5′ Position	3′ Position
13	3272350CB1	1600	1573-1600, 1-53, 433-614, 1505-1542	GNN.g8517773_1.edit	1	664
				2013147H1 (TESTNOT03)	1431	1600
				8094106H1 (EYERNOA01)	473	1133
				71163473V1	1099	1593
14	7481507CB1	781	745-781, 395-691, 1-313	GBI.g7188861_000153.edit	426	781
				GNN.g6446924_004.edit	80	644
				55001533J2	1	307
15	2285140CB1	1724	871-914, 1-52, 1296-1724	3271918H1 (BRAINOT20)	1	239
				362853R6 (PROSNOT01)	194	756
				362853T6 (PROSNOT01)	633	1299
				1856725F6 (PROSNOT18)	1110	1724
				2173313T6 (ENDCNOT03)	317	878
16	7197873CB1	4157	1-56, 3888-3949, 741-2915	55099335H1	1	835
				70880928V1	3513	4157
				72010790V1	3152	3828
				55075261J1	2247	2848
				55123062J1	498	1292
				55099328J1	2518	3215
				56000513H1	1238	1974
				72008877V1	3206	3950
				55076893J1	1877	2420
17	6282188CB1	1044	1-1044	71715772V1	1	678
				71715368V1	414	1044
18	2182961CB1	2797	1394-2240, 1-77, 285-803	58002040T1	2151	2797
				55144256J1	1155	2046
				2893561H1 (KIDNTUT14)	1	286
				114545H1 (TESTNOT01)	846	971
				GNN:g8570194	213	1475
				58002164T1	2010	2794
19	5119906CB1	3488	2861-3488, 1-1396	g4533101	1794	2204
				5079017F6 (LNODNOT11)	1873	2064
				GNN.g6978120_000001_002	874	3488
				6814714J1 (ADRETUR01)	1	876
				5119906F6 (SMCBUNT01)	946	1488
				7441274R6 (ADRETUE02)	1521	2005
				g1997526	2139	2232
				6565672H1 (MCLDTXT04)	808	1427

Table 4 (cont.)

Polynucleotide SEQ ID NO:	Incyte Polynucleotide ID	Sequence Length	Selected Fragment(s)	Sequence Fragments	5' Position	3' Position
20	4022502CB1	1522	1-109, 1425-1522	g2012311	1372	1522
				5594812H1 (COLCDIT03)	1243	1505
				7947957J1 (BRABNOE02)	45	743
				3540664H1 (SEMVNOT04)	1339	1517
				4022502F8 (BRAXNOT02)	396	1058
				5594812F6 (COLCDIT03)	998	1502
				6299318H1 (UTREDIT07)	1	283
21	4084356CB1	1393	1366-1393, 1-47, 706-807	6332847H1 (BRANDIN01)	597	1109
				GNN.g809120_006.edit	106	1095
				7333518H1 (CONFTDN02)	917	1393
				GBI.g809120_000001.edit	1	810
22	1740204CB1	1430	1-401	g3163696	764	1107
				6332536H1 (BRANDIN01)	1034	1430
				6205996H1 (PITUNON01)	204	912
				70218058V1	1	508
23	7483804CB1	3102	1-990, 1441-1460, 1840-2628, 3083-3102	7189648H1 (BRATDIC01)	1	500
				6873131H1 (BRAGNON02)	2207	2941
				72470166D1	1438	2223
				72475127D1	2273	3055
				72474804D1	615	1302
				72474643D1	2315	3102
				71880642V1	1270	2110
				6987688R8 (BRAIFER05)	328	1243
24	7483934CB1	5612	3420-3503, 1-313, 824-2655, 4533-4960	8114162H1 (OSTEUNC01)	4977	5612
				8130776H1 (SCOMDIC01)	4420	5121
				7483934CB1	59	5063
				4672080H1 (SINTNOT24)	1	223

Table 5

Polynucleotide SEQ ID NO:	Incyte Project ID	Representative Library
13	3272350CB1	OVARNOT13
15	2285140CB1	BRSTNOT01
16	7197873CB1	BRAINOT12
17	6282188CB1	SKINDIA01
18	2182961CB1	SININOT01
19	5119906CB1	SMCBUNT01
20	4022502CB1	BRAXNOT02
21	4084356CB1	CONFNOT02
22	1740204CB1	BRAINOT09
23	7483804CB1	BSCNNOT03
24	7483934CB1	BRAUNOR01

Table 6

Library	Vector	Library Description
BRAINOT09	pINCY	Library was constructed using RNA isolated from brain tissue removed from a Caucasian male fetus, who died at 23 weeks' gestation.
BRAINOT12	pINCY	Library was constructed using RNA isolated from brain tissue removed from the right frontal lobe of a 5-year-old Caucasian male during a hemispherectomy. Pathology indicated extensive polymicrogyria and mild to moderate gliosis (predominantly subpial and subcortical), which are consistent with chronic seizure disorder. Family history included a cervical neoplasm.
BRAUNOR01	pINCY	This random primed library was constructed using RNA isolated from striatum, globus pallidus and posterior putamen tissue removed from an 81-year-old Caucasian female who died from a hemorrhage and ruptured thoracic aorta due to atherosclerosis. Pathology indicated moderate atherosclerosis involving the internal carotids bilaterally; microscopic infarcts of the frontal cortex and hippocampus; and scattered diffuse amyloid plaques and neurofibrillary tangles, consistent with age. Grossly, the leptomeninges showed only mild thickening and hyalinization along the superior sagittal sinus. The remainder of the leptomeninges was thin and contained some congested blood vessels. Mild atrophy was found mostly in the frontal poles and lobes, and temporal lobes, bilaterally. Microscopically, there were pairs of Alzheimer type II astrocytes within the deep layers of the neocortex. There was increased satellitosis around neurons in the deep gray matter in the middle frontal cortex. The amygdala contained rare diffuse plaques and neurofibrillary tangles. The posterior hippocampus contained a microscopic area of cystic cavitation with hemosiderin-laden macrophages surrounded by reactive gliosis. Patient history included sepsis, cholangitis, post-operative atelectasis, pneumonia CAD, cardiomegaly due to left ventricular hypertrophy, splenomegaly, arteriolonephrosclerosis, nodular colloidal goiter, emphysema, CHF, hypothyroidism, and peripheral vascular disease.
BRAXNOT02	pINCY	Library was constructed using RNA isolated from cerebellar tissue removed from a 64-year-old male. Patient history included carcinoma of the left bronchus.
BRSTNOT01	PBLUE-SCRIPT	Library was constructed using RNA isolated from the breast tissue of a 56-year-old Caucasian female who died in a motor vehicle accident.
BSCNNOT03	pINCY	Library was constructed using RNA isolated from caudate nucleus tissue removed from the brain of a 92-year-old male. Pathology indicated several small cerebral infarcts but no senile plaques or neurofibrillary degeneration. Patient history included throat cancer which was treated with radiation.
CONFNOT02	pINCY	Library was constructed using RNA isolated from abdominal fat tissue removed from a 52-year-old Caucasian female during an ileum resection and incarcerated ventral hernia repair. Patient history included diverticulitis. Family history included hyperlipidemia.

Table 6 (cont.)

Library	Vector	Library Description
OVARNOT13	pINCY	Library was constructed using RNA isolated from left ovary tissue removed from a 47-year-old Caucasian female during a vaginal hysterectomy with bilateral salpingo-oophorectomy, and dilation and curettage. Pathology for the associated tumor tissue indicated a single intramural leiomyoma. The endometrium was in the secretory phase. The patient presented with metrorrhagia. Patient history included hyperlipidemia and benign hypertension. Family history included colon cancer, benign hypertension, atherosclerotic coronary artery disease, and breast cancer.
SININOT01	pINCY	Library was constructed using RNA isolated from ileum tissue obtained from the small intestine of a 4-year-old Caucasian female, who died from a closed head injury. Serologies were negative. Patient history included jaundice. Previous surgeries included a double hernia repair.
SKINDIA01	PSPORT1	This amplified library was constructed using RNA isolated from diseased skin tissue removed from 1 female and 4 males during skin biopsies. Pathologies indicated tuberculoid and lepromatious leprosy.
SMCBUNT01	pINCY	Library was constructed using RNA isolated from untreated bronchial smooth muscle cell tissue removed from a 21-year-old Caucasian male.

Table 7

Program	Description	Reference	Parameter Threshold
ABI FACTURA	A program that removes vector sequences and masks ambiguous bases in nucleic acid sequences.	Applied Biosystems, Foster City, CA.	
ABI/PARACEL FDF	A Fast Data Finder useful in comparing and annotating amino acid or nucleic acid sequences.	Applied Biosystems, Foster City, CA; Paracel Inc., Pasadena, CA.	Mismatch <50%
ABI AutoAssembler	A program that assembles nucleic acid sequences.	Applied Biosystems, Foster City, CA.	
BLAST	A Basic Local Alignment Search Tool useful in sequence similarity search for amino acid and nucleic acid sequences. BLAST includes five functions: blastp, blastn, blastx, tblastn, and tblastx.	Altschul, S.F. et al. (1990) J. Mol. Biol. 215:403-410; Altschul, S.F. et al. (1997) Nucleic Acids Res. 25:3389-3402.	*ESTs:* Probability value= 1.0E-8 or less *Full Length sequences*: Probability value= 1.0E-10 or less
FASTA	A Pearson and Lipman algorithm that searches for similarity between a query sequence and a group of sequences of the same type. FASTA comprises as least five functions: fasta, tfasta, fastx, tfastx, and ssearch.	Pearson, W.R. and D.J. Lipman (1988) Proc. Natl. Acad Sci. USA 85:2444-2448; Pearson, W.R. (1990) Methods Enzymol. 183:63-98; and Smith, T.F. and M.S. Waterman (1981) Adv. Appl. Math. 2:482-489.	*ESTs:* fasta E value=1.06E-6 *Assembled ESTs*: fasta Identity= 95% or greater and Match length=200 bases or greater; fastx E value=1.0E-8 or less *Full Length sequences:* fastx score=100 or greater
BLIMPS	A BLocks IMProved Searcher that matches a sequence against those in BLOCKS, PRINTS, DOMO, PRODOM, and PFAM databases to search for gene families, sequence homology, and structural fingerprint regions.	Henikoff, S. and J.G. Henikoff (1991) Nucleic Acids Res. 19:6565-6572; Henikoff, J.G. and S. Henikoff (1996) Methods Enzymol. 266:88-105; and Attwood, T.K. et al. (1997) J. Chem. Inf. Comput. Sci. 37:417-424.	Probability value= 1.0E-3 or less
HMMER	An algorithm for searching a query sequence against hidden Markov model (HMM)-based databases of protein family consensus sequences, such as PFAM.	Krogh, A. et al. (1994) J. Mol. Biol. 235:1501-1531; Sonnhammer, E.L.L. et al. (1988) Nucleic Acids Res. 26:320-322; Durbin, R. et al. (1998) Our World View, in a Nutshell, Cambridge Univ. Press, pp. 1-350.	*PFAM hits:* Probability value= 1.0E-3 or less *Signal peptide hits:* Score= 0 or greater

Table 7 (cont.)

Program	Description	Reference	Parameter Threshold
ProfileScan	An algorithm that searches for structural and sequence motifs in protein sequences that match sequence patterns defined in Prosite.	Gribskov, M. et al. (1988) CABIOS 4:61-66; Gribskov, M. et al. (1989) Methods Enzymol. 183:146-159; Bairoch, A. et al. (1997) Nucleic Acids Res. 25:217-221.	Normalized quality score≥GCG-specified "HIGH" value for that particular Prosite motif. Generally, score=1.4-2.1.
Phred	A base-calling algorithm that examines automated sequencer traces with high sensitivity and probability.	Ewing, B. et al. (1998) Genome Res. 8:175-185; Ewing, B. and P. Green (1998) Genome Res. 8:186-194.	
Phrap	A Phils Revised Assembly Program including SWAT and CrossMatch, programs based on efficient implementation of the Smith-Waterman algorithm, useful in searching sequence homology and assembling DNA sequences.	Smith, T.F. and M.S. Waterman (1981) Adv. Appl. Math. 2:482-489; Smith, T.F. and M.S. Waterman (1981) J. Mol. Biol. 147:195-197; and Green, P., University of Washington, Seattle, WA.	Score= 120 or greater; Match length= 56 or greater
Consed	A graphical tool for viewing and editing Phrap assemblies.	Gordon, D. et al. (1998) Genome Res. 8:195-202.	
SPScan	A weight matrix analysis program that scans protein sequences for the presence of secretory signal peptides.	Nielson, H. et al. (1997) Protein Engineering 10:1-6; Claverie, J.M. and S. Audic (1997) CABIOS 12:431-439.	Score=3.5 or greater
TMAP	A program that uses weight matrices to delineate transmembrane segments on protein sequences and determine orientation.	Persson, B. and P. Argos (1994) J. Mol. Biol. 237:182-192; Persson, B. and P. Argos (1996) Protein Sci. 5:363-371.	
TMHMMER	A program that uses a hidden Markov model (HMM) to delineate transmembrane segments on protein sequences and determine orientation.	Sonnhammer, E.L. et al. (1998) Proc. Sixth Intl. Conf. on Intelligent Systems for Mol. Biol., Glasgow et al., eds., The Am. Assoc. for Artificial Intelligence Press, Menlo Park, CA, pp. 175-182.	
Motifs	A program that searches amino acid sequences for patterns that matched those defined in Prosite.	Bairoch, A. et al. (1997) Nucleic Acids Res. 25:217-221; Wisconsin Package Program Manual, version 9, page M51-59, Genetics Computer Group, Madison, WI.	

What is claimed is:

1. An isolated polypeptide selected from the group consisting of:
   - a) a polypeptide comprising an amino acid sequence selected from the group consisting of SEQ ID NO:1-12,
   - b) a polypeptide comprising a naturally occurring amino acid sequence at least 90% identical to an amino acid sequence selected from the group consisting of SEQ ID NO:1-12,
   - c) a biologically active fragment of a polypeptide having an amino acid sequence selected from the group consisting of SEQ ID NO:1-12, and
   - d) an immunogenic fragment of a polypeptide having an amino acid sequence selected from the group consisting of SEQ ID NO:1-12.

2. An isolated polypeptide of claim 1 selected from the group consisting of SEQ ID NO:1-12.

3. An isolated polynucleotide encoding a polypeptide of claim 1.

4. An isolated polynucleotide encoding a polypeptide of claim 2.

5. An isolated polynucleotide of claim 4 selected from the group consisting of SEQ ID NO:13-24.

6. A recombinant polynucleotide comprising a promoter sequence operably linked to a polynucleotide of claim 3.

7. A cell transformed with a recombinant polynucleotide of claim 6.

8. A transgenic organism comprising a recombinant polynucleotide of claim 6.

9. A method of producing a polypeptide of claim 1, the method comprising:
   - a) culturing a cell under conditions suitable for expression of the polypeptide, wherein said cell is transformed with a recombinant polynucleotide, and said recombinant polynucleotide comprises a promoter sequence operably linked to a polynucleotide encoding the polypeptide of claim 1, and
   - b) recovering the polypeptide so expressed.

10. A method of claim 9, wherein the polypeptide has an amino acid sequence selected from the group consisting of SEQ ID NO:1-12.

11. An isolated antibody which specifically binds to a polypeptide of claim 1.

12. An isolated polynucleotide selected from the group consisting of:

a) a polynucleotide comprising a polynucleotide sequence selected from the group consisting of SEQ ID NO:13-24,

b) a polynucleotide comprising a naturally occurring polynucleotide sequence at least 90% identical to a polynucleotide sequence selected from the group consisting of SEQ ID NO:13-24,

c) a polynucleotide complementary to a polynucleotide of a),

d) a polynucleotide complementary to a polynucleotide of b), and

e) an RNA equivalent of a)-d).

13. An isolated polynucleotide comprising at least 60 contiguous nucleotides of a polynucleotide of claim 12.

14. A method of detecting a target polynucleotide in a sample, said target polynucleotide having a sequence of a polynucleotide of claim 12, the method comprising:

a) hybridizing the sample with a probe comprising at least 20 contiguous nucleotides comprising a sequence complementary to said target polynucleotide in the sample, and which probe specifically hybridizes to said target polynucleotide, under conditions whereby a hybridization complex is formed between said probe and said target polynucleotide or fragments thereof, and

b) detecting the presence or absence of said hybridization complex, and, optionally, if present, the amount thereof.

15. A method of claim 14, wherein the probe comprises at least 60 contiguous nucleotides.

16. A method of detecting a target polynucleotide in a sample, said target polynucleotide having a sequence of a polynucleotide of claim 12, the method comprising:

a) amplifying said target polynucleotide or fragment thereof using polymerase chain reaction amplification, and

b) detecting the presence or absence of said amplified target polynucleotide or fragment thereof, and, optionally, if present, the amount thereof.

17. A composition comprising a polypeptide of claim 1 and a pharmaceutically acceptable excipient.

18. A composition of claim 17, wherein the polypeptide has an amino acid sequence selected from the group consisting of SEQ ID NO:1-12.

19. A method for treating a disease or condition associated with decreased expression of functional PP, comprising administering to a patient in need of such treatment the composition of claim 17.

20. A method of screening a compound for effectiveness as an agonist of a polypeptide of claim 1, the method comprising:

a) exposing a sample comprising a polypeptide of claim 1 to a compound, and

b) detecting agonist activity in the sample.

21. A composition comprising an agonist compound identified by a method of claim 20 and a pharmaceutically acceptable excipient.

22. A method for treating a disease or condition associated with decreased expression of functional PP, comprising administering to a patient in need of such treatment a composition of claim 21.

23. A method of screening a compound for effectiveness as an antagonist of a polypeptide of claim 1, the method comprising:

a) exposing a sample comprising a polypeptide of claim 1 to a compound, and

b) detecting antagonist activity in the sample.

24. A composition comprising an antagonist compound identified by a method of claim 23 and a pharmaceutically acceptable excipient.

25. A method for treating a disease or condition associated with overexpression of functional PP, comprising administering to a patient in need of such treatment a composition of claim 24.

26. A method of screening for a compound that specifically binds to the polypeptide of claim 1, the method comprising:

a) combining the polypeptide of claim 1 with at least one test compound under suitable conditions, and

b) detecting binding of the polypeptide of claim 1 to the test compound, thereby identifying a compound that specifically binds to the polypeptide of claim 1.

27. A method of screening for a compound that modulates the activity of the polypeptide of claim 1, the method comprising:

a) combining the polypeptide of claim 1 with at least one test compound under conditions permissive for the activity of the polypeptide of claim 1,

b) assessing the activity of the polypeptide of claim 1 in the presence of the test compound, and

c) comparing the activity of the polypeptide of claim 1 in the presence of the test compound with the activity of the polypeptide of claim 1 in the absence of the test compound, wherein a change in the activity of the polypeptide of claim 1 in the presence of the test compound is indicative of a compound that modulates the activity of the polypeptide of claim 1.

28. A method of screening a compound for effectiveness in altering expression of a target polynucleotide, wherein said target polynucleotide comprises a sequence of claim 5, the method comprising:

a) exposing a sample comprising the target polynucleotide to a compound, under conditions suitable for the expression of the target polynucleotide,

b) detecting altered expression of the target polynucleotide, and

c) comparing the expression of the target polynucleotide in the presence of varying amounts of the compound and in the absence of the compound.

29. A method of assessing toxicity of a test compound, the method comprising:

a) treating a biological sample containing nucleic acids with the test compound,

b) hybridizing the nucleic acids of the treated biological sample with a probe comprising at least 20 contiguous nucleotides of a polynucleotide of claim 12 under conditions whereby a specific hybridization complex is formed between said probe and a target polynucleotide in the biological sample, said target polynucleotide comprising a polynucleotide sequence of a polynucleotide of claim 12 or fragment thereof,

c) quantifying the amount of hybridization complex, and

d) comparing the amount of hybridization complex in the treated biological sample with the amount of hybridization complex in an untreated biological sample, wherein a difference in the amount of hybridization complex in the treated biological sample is indicative of toxicity of the test compound.

30. A diagnostic test for a condition or disease associated with the expression of PP in a biological sample, the method comprising:

a) combining the biological sample with an antibody of claim 11, under conditions suitable for the antibody to bind the polypeptide and form an antibody:polypeptide complex, and

b) detecting the complex, wherein the presence of the complex correlates with the presence of the polypeptide in the biological sample.

31. The antibody of claim 11, wherein the antibody is:

a) a chimeric antibody,

b) a single chain antibody,

c) a Fab fragment,

d) a $F(ab')_2$ fragment, or

e) a humanized antibody.

32. A composition comprising an antibody of claim 11 and an acceptable excipient.

33. A method of diagnosing a condition or disease associated with the expression of PP in a subject, comprising administering to said subject an effective amount of the composition of claim 32.

34. A composition of claim 32, wherein the antibody is labeled.

35. A method of diagnosing a condition or disease associated with the expression of PP in a subject, comprising administering to said subject an effective amount of the composition of claim 34.

36. A method of preparing a polyclonal antibody with the specificity of the antibody of claim 11, the method comprising:

a) immunizing an animal with a polypeptide having an amino acid sequence selected from the group consisting of SEQ ID NO:1-12, or an immunogenic fragment thereof, under conditions to elicit an antibody response,

b) isolating antibodies from said animal, and

c) screening the isolated antibodies with the polypeptide, thereby identifying a polyclonal antibody which binds specifically to a polypeptide having an amino acid sequence selected from the group consisting of SEQ ID NO:1-12.

37. A polyclonal antibody produced by a method of claim 36.

38. A composition comprising the polyclonal antibody of claim 37 and a suitable carrier.

39. A method of making a monoclonal antibody with the specificity of the antibody of claim 11, the method comprising:

a) immunizing an animal with a polypeptide having an amino acid sequence selected from the group consisting of SEQ ID NO:1-12, or an immunogenic fragment thereof, under conditions to elicit an antibody response,

b) isolating antibody producing cells from the animal,

c) fusing the antibody producing cells with immortalized cells to form monoclonal antibody-producing hybridoma cells,

d) culturing the hybridoma cells, and

e) isolating from the culture monoclonal antibody which binds specifically to a polypeptide having an amino acid sequence selected from the group consisting of SEQ ID NO:1-12.

40. A monoclonal antibody produced by a method of claim 39.

41. A composition comprising the monoclonal antibody of claim 40 and a suitable carrier.

42. The antibody of claim 11, wherein the antibody is produced by screening a Fab expression library.

43. The antibody of claim 11, wherein the antibody is produced by screening a recombinant immunoglobulin library.

44. A method of detecting a polypeptide having an amino acid sequence selected from the group consisting of SEQ ID NO:1-12 in a sample, the method comprising:

a) incubating the antibody of claim 11 with a sample under conditions to allow specific binding of the antibody and the polypeptide, and
b) detecting specific binding, wherein specific binding indicates the presence of a polypeptide having an amino acid sequence selected from the group consisting of SEQ ID NO:1-12 in the sample.

45. A method of purifying a polypeptide having an amino acid sequence selected from the group consisting of SEQ ID NO:1-12 from a sample, the method comprising:

a) incubating the antibody of claim 11 with a sample under conditions to allow specific binding of the antibody and the polypeptide, and
b) separating the antibody from the sample and obtaining the purified polypeptide having an amino acid sequence selected from the group consisting of SEQ ID NO:1-12.

46. A microarray wherein at least one element of the microarray is a polynucleotide of claim 13.

47. A method of generating a transcript image of a sample which contains polynucleotides, the method comprising:

a) labeling the polynucleotides of the sample,
b) contacting the elements of the microarray of claim 46 with the labeled polynucleotides of the sample under conditions suitable for the formation of a hybridization complex, and
c) quantifying the expression of the polynucleotides in the sample.

48. An array comprising different nucleotide molecules affixed in distinct physical locations on a solid substrate, wherein at least one of said nucleotide molecules comprises a first oligonucleotide or polynucleotide sequence specifically hybridizable with at least 30 contiguous nucleotides of a target polynucleotide, and wherein said target polynucleotide is a polynucleotide of claim 12.

49. An array of claim 48, wherein said first oligonucleotide or polynucleotide sequence is completely complementary to at least 30 contiguous nucleotides of said target polynucleotide.

50. An array of claim 48, wherein said first oligonucleotide or polynucleotide sequence is completely complementary to at least 60 contiguous nucleotides of said target polynucleotide.

51. An array of claim 48, wherein said first oligonucleotide or polynucleotide sequence is completely complementary to said target polynucleotide.

52. An array of claim 48, which is a microarray.

53. An array of claim 48, further comprising said target polynucleotide hybridized to a nucleotide molecule comprising said first oligonucleotide or polynucleotide sequence.

54. An array of claim 48, wherein a linker joins at least one of said nucleotide molecules to said solid substrate.

55. An array of claim 48, wherein each distinct physical location on the substrate contains multiple nucleotide molecules, and the multiple nucleotide molecules at any single distinct physical location have the same sequence, and each distinct physical location on the substrate contains nucleotide molecules having a sequence which differs from the sequence of nucleotide molecules at another distinct physical location on the substrate.

56. A polypeptide of claim 1, comprising the amino acid sequence of SEQ ID NO:1.

57. A polypeptide of claim 1, comprising the amino acid sequence of SEQ ID NO:2.

58. A polypeptide of claim 1, comprising the amino acid sequence of SEQ ID NO:3.

59. A polypeptide of claim 1, comprising the amino acid sequence of SEQ ID NO:4.

60. A polypeptide of claim 1, comprising the amino acid sequence of SEQ ID NO:5.

61. A polypeptide of claim 1, comprising the amino acid sequence of SEQ ID NO:6.

62. A polypeptide of claim 1, comprising the amino acid sequence of SEQ ID NO:7.

63. A polypeptide of claim 1, comprising the amino acid sequence of SEQ ID NO:8.

64. A polypeptide of claim 1, comprising the amino acid sequence of SEQ ID NO:9.

65. A polypeptide of claim 1, comprising the amino acid sequence of SEQ ID NO:10.

66. A polypeptide of claim 1, comprising the amino acid sequence of SEQ ID NO:11.

67. A polypeptide of claim 1, comprising the amino acid sequence of SEQ ID NO:12.

68. A polynucleotide of claim 12, comprising the polynucleotide sequence of SEQ ID NO:13.

69. A polynucleotide of claim 12, comprising the polynucleotide sequence of SEQ ID NO:14.

70. A polynucleotide of claim 12, comprising the polynucleotide sequence of SEQ ID NO:15.

71. A polynucleotide of claim 12, comprising the polynucleotide sequence of SEQ ID NO:16.

72. A polynucleotide of claim 12, comprising the polynucleotide sequence of SEQ ID NO:17.

73. A polynucleotide of claim 12, comprising the polynucleotide sequence of SEQ ID NO:18.

74. A polynucleotide of claim 12, comprising the polynucleotide sequence of SEQ ID NO:19.

75. A polynucleotide of claim 12, comprising the polynucleotide sequence of SEQ ID NO:20.

76. A polynucleotide of claim 12, comprising the polynucleotide sequence of SEQ ID NO:21.

77. A polynucleotide of claim 12, comprising the polynucleotide sequence of SEQ ID NO:22.

78. A polynucleotide of claim 12, comprising the polynucleotide sequence of SEQ ID NO:23.

79. A polynucleotide of claim 12, comprising the polynucleotide sequence of SEQ ID NO:24.

<110> INCYTE GENOMICS, INC.
TANG, Y. Tom
YAO, Monique G.
WALIA, Narinder K.
ELLIOTT, Vicki S.
RAMKUMAR, Jayalaxmi
LU, Yan
ARVIZU, Chandra
DING, Li
BAUGHN, Mariah R.
YUE, Henry
LU, Dyung Aina M.
TRIBOULEY, Catherine M.
THORNTON, Michael
GANDHI, Ameena R.
LEE, Ernestine A.
XU, Yuming
WANG, Yumei E.
HAFALIA, April J.A.
THANGAVELU, Kavitha
DANIELS, Susan E.
LAL, Preeti
SWARNAKAR, Anita

<120> PROTEIN PHOSPHATASES

<130> PI-0228 PCT

<140> To Be Assigned
<141> Herewith

<150> 60/234,526; 60/236,967; 60/238,332; 60/242,236; 60/243,928; 60/249,814
<151> 2000-09-21; 2000-09-29; 2000-10-05; 2000-10-19; 2000-10-27; 2000-11-17

<160> 24

<170> PERL Program

<210> 1
<211> 435
<212> PRT
<213> Homo sapiens

<220>
<221> misc_feature
<223> Incyte ID No: 3272350CD1

<400> 1

```
Met Ala Gly Glu Asn Gly Gln Glu Gly Val Gly Ile Cys Arg Leu
 1               5                  10                  15
Gly Val Gln Pro Glu Val Glu Pro Ser Ser Gln Asp Val Arg Gln
                20                  25                  30
Ala Leu Gly Arg Pro Val Leu Leu Arg Cys Ser Leu Leu Arg Gly
                35                  40                  45
Ser Pro Gln Arg Ile Ala Ser Ala Val Trp Arg Phe Lys Gly Gln
                50                  55                  60
Leu Leu Pro Pro Pro Pro Val Val Pro Ala Ala Ala Glu Ala Pro
                65                  70                  75
Asp His Ala Glu Leu Arg Leu Asp Ala Val Thr Arg Asp Ser Ser
                80                  85                  90
Gly Ser Tyr Glu Cys Ser Val Ser Asn Asp Val Gly Ser Ala Ala
                95                 100                 105
```

```
Cys Leu Phe Gln Val Ser Ala Lys Ala Tyr Ser Pro Glu Phe Tyr
                110                 115                 120
Phe Asp Thr Pro Asn Pro Thr Arg Ser His Lys Leu Ser Lys Asn
                125                 130                 135
Tyr Ser Tyr Val Leu Gln Trp Thr Gln Arg Glu Pro Asp Ala Val
                140                 145                 150
Asp Pro Val Leu Asn Tyr Arg Leu Ser Ile Arg Gln Leu Asn Gln
                155                 160                 165
His Asn Ala Val Val Lys Ala Ile Pro Val Arg Arg Val Glu Lys
                170                 175                 180
Gly Gln Leu Leu Glu Tyr Ile Leu Thr Asp Leu Arg Val Pro His
                185                 190                 195
Ser Tyr Glu Val Arg Leu Thr Pro Tyr Thr Thr Phe Gly Ala Gly
                200                 205                 210
Asp Met Ala Ser Arg Ile Ile His Tyr Thr Glu Pro Ile Asn Ser
                215                 220                 225
Pro Asn Leu Ser Asp Asn Thr Cys His Phe Glu Asp Glu Lys Ile
                230                 235                 240
Cys Gly Tyr Thr Gln Asp Leu Thr Asp Asn Phe Asp Trp Thr Arg
                245                 250                 255
Gln Asn Ala Leu Thr Gln Asn Pro Lys Arg Ser Pro Asn Thr Gly
                260                 265                 270
Pro Pro Thr Asp Ile Ser Gly Thr Pro Glu Gly Tyr Tyr Met Phe
                275                 280                 285
Ile Glu Thr Ser Arg Pro Arg Glu Leu Gly Asp Arg Ala Arg Leu
                290                 295                 300
Val Ser Pro Leu Tyr Asn Ala Ser Ala Lys Phe Tyr Cys Val Ser
                305                 310                 315
Phe Phe Tyr His Met Tyr Gly Lys His Ile Gly Ser Leu Asn Leu
                320                 325                 330
Leu Val Arg Ser Arg Asn Lys Gly Ala Leu Asp Thr His Ala Trp
                335                 340                 345
Ser Leu Ser Gly Asn Lys Gly Asn Val Trp Gln Gln Ala His Val
                350                 355                 360
Pro Ile Ser Pro Ser Gly Pro Phe Gln Ile Ile Phe Glu Gly Val
                365                 370                 375
Arg Gly Pro Gly Tyr Leu Gly Asp Ile Ala Ile Asp Asp Val Thr
                380                 385                 390
Leu Lys Lys Gly Glu Cys Pro Arg Lys Gln Thr Asp Pro Asn Lys
                395                 400                 405
Val Val Val Met Pro Gly Ser Gly Ala Pro Cys Gln Ser Ser Pro
                410                 415                 420
Gln Leu Trp Gly Pro Met Ala Ile Phe Leu Leu Ala Leu Gln Arg
                425                 430                 435


<210> 2
<211> 233
<212> PRT
<213> Homo sapiens

<220>
<221> misc_feature
<223> Incyte ID No: 7481507CD1

<400> 2
Met Ala Ala Thr Arg Gly Glu Glu Lys Ile Cys Met Ser Met Tyr
  1               5                  10                  15
Gln Arg Ile Asn Gly Ala Asp Trp Arg Asn Ile Phe Val Val Gly
                 20                  25                  30
Asp Leu His Gly Cys Tyr Thr Leu Leu Met Asn Glu Leu Glu Lys
                 35                  40                  45
Val Ser Phe Asp Pro Ala Cys Asp Leu Leu Ile Ser Val Gly Asp
```

```
                50                  55                  60
Leu Val Asp Arg Gly Ala Glu Asn Val Glu Cys Leu Glu Leu Ile
                65                  70                  75
Thr Met Pro Trp Phe Arg Ala Val Arg Gly Asn His Glu Gln Met
                80                  85                  90
Met Ile Asp Gly Leu Ser Glu Tyr Gly Asn Val Asn His Trp Leu
                95                 100                 105
Glu Asn Gly Gly Val Trp Phe Phe Ser Leu Asp Tyr Glu Lys Glu
               110                 115                 120
Val Leu Ala Lys Ala Leu Val His Lys Ser Ala Ser Leu Pro Phe
               125                 130                 135
Val Ile Glu Leu Val Thr Ala Glu Arg Lys Ile Val Ile Cys His
               140                 145                 150
Ala Asp Tyr Pro His Asn Glu Tyr Ala Phe Asp Lys Pro Val Pro
               155                 160                 165
Lys Asp Met Val Ile Trp Asn Arg Glu Arg Val Ser Asp Ala Gln
               170                 175                 180
Asp Gly Ile Val Ser Pro Ile Ala Gly Ala Asp Leu Phe Ile Phe
               185                 190                 195
Gly His Thr Pro Ala Arg Gln Pro Leu Lys Tyr Ala Asn Gln Met
               200                 205                 210
Tyr Ile Asp Thr Gly Ala Val Phe Cys Gly Asn Leu Thr Leu Val
               215                 220                 225
Gln Val Gln Gly Gly Ala His Ala
               230

<210> 3
<211> 315
<212> PRT
<213> Homo sapiens

<220>
<221> misc_feature
<223> Incyte ID No: 2285140CD1

<400> 3
Met His Gly His Gly Gly Tyr Asp Ser Asp Phe Ser Asp Asp Glu
 1               5                  10                  15
His Cys Gly Glu Ser Ser Lys Arg Lys Lys Arg Thr Val Glu Asp
                20                  25                  30
Asp Leu Leu Leu Gln Lys Pro Phe Gln Lys Glu Lys His Gly Lys
                35                  40                  45
Val Ala His Lys Gln Val Ala Ala Glu Leu Leu Asp Arg Glu Glu
                50                  55                  60
Ala Arg Asn Arg Arg Phe His Leu Ile Ala Met Asp Ala Tyr Gln
                65                  70                  75
Arg His Arg Lys Phe Val Asn Asp Tyr Ile Leu Tyr Tyr Gly Gly
                80                  85                  90
Lys Lys Glu Asp Phe Lys Arg Leu Gly Glu Asn Asp Lys Thr Asp
                95                 100                 105
Leu Asp Val Ile Arg Glu Asn His Arg Phe Leu Trp Asn Glu Glu
               110                 115                 120
Asp Glu Met Asp Met Thr Trp Glu Lys Arg Leu Ala Lys Lys Tyr
               125                 130                 135
Tyr Asp Lys Leu Phe Lys Glu Tyr Cys Ile Ala Asp Leu Ser Lys
               140                 145                 150
Tyr Lys Glu Asn Lys Phe Gly Phe Arg Trp Arg Val Glu Lys Glu
               155                 160                 165
Val Ile Ser Gly Lys Gly Gln Phe Phe Cys Gly Asn Lys Tyr Cys
               170                 175                 180
Asp Lys Lys Glu Gly Leu Lys Ser Trp Glu Val Asn Phe Gly Tyr
               185                 190                 195
Ile Glu His Gly Glu Lys Arg Asn Ala Leu Val Lys Leu Arg Leu
```

```
                200                 205                 210
Cys Gln Glu Cys Ser Ile Lys Leu Asn Phe His His Arg Arg Lys
                215                 220                 225
Glu Ile Lys Ser Lys Lys Arg Lys Asp Lys Thr Lys Lys Asp Cys
                230                 235                 240
Glu Glu Ser Ser His Lys Lys Ser Arg Leu Ser Ser Ala Glu Glu
                245                 250                 255
Ala Ser Lys Lys Lys Asp Lys Gly His Ser Ser Ser Lys Lys Ser
                260                 265                 270
Glu Asp Ser Leu Leu Arg Asn Ser Asp Glu Glu Glu Ser Ala Ser
                275                 280                 285
Glu Ser Glu Leu Trp Lys Gly Pro Leu Pro Glu Thr Asp Glu Lys
                290                 295                 300
Ser Gln Glu Glu Glu Phe Asp Glu Tyr Phe Gln Asp Leu Phe Leu
                305                 310                 315


<210> 4
<211> 1278
<212> PRT
<213> Homo sapiens

<220>
<221> misc_feature
<223> Incyte ID No: 7197873CD1

<400> 4
Met Ser Leu Ser Ser Val Thr Leu Ala Ser Ala Leu Gln Val Arg
 1                5                  10                  15
Gly Glu Ala Leu Ser Glu Glu Glu Ile Trp Ser Pro Leu Phe Leu
                 20                  25                  30
Ala Ala Glu Gln Leu Leu Glu Asp Leu Arg Asn Asp Ser Ser Asp
                 35                  40                  45
Tyr Val Val Cys Pro Trp Ser Ala Leu Leu Ser Ala Ala Gly Ser
                 50                  55                  60
Leu Ser Phe Gln Gly Arg Val Ser His Ile Glu Ala Ala Pro Phe
                 65                  70                  75
Lys Ala Pro Glu Leu Leu Gln Gly Gln Ser Glu Asp Glu Gln Pro
                 80                  85                  90
Asp Ala Ser Gln Pro Leu Gln Leu Cys Glu Pro Leu His Ser Ile
                 95                 100                 105
Leu Leu Thr Met Cys Glu Asp Gln Pro His Arg Arg Cys Thr Leu
                110                 115                 120
Gln Ser Val Leu Glu Ala Cys Arg Val His Glu Lys Glu Val Ser
                125                 130                 135
Val Tyr Pro Ala Pro Ala Gly Leu His Ile Arg Arg Leu Val Gly
                140                 145                 150
Leu Val Leu Gly Thr Ile Ser Glu Val Glu Lys Arg Val Val Glu
                155                 160                 165
Glu Ser Ser Ser Val Gln Gln Asn Arg Ser Tyr Leu Leu Arg Lys
                170                 175                 180
Arg Leu Arg Gly Thr Ser Ser Glu Ser Pro Ala Ala Gln Ala Pro
                185                 190                 195
Glu Cys Leu His Pro Cys Arg Val Ser Glu Arg Ser Thr Glu Thr
                200                 205                 210
Gln Ser Ser Pro Glu Pro His Trp Ser Thr Leu Thr His Ser His
                215                 220                 225
Cys Ser Leu Leu Val Asn Arg Ala Leu Pro Gly Ala Asp Pro Gln
                230                 235                 240
Asp Gln Gln Ala Gly Arg Arg Leu Ser Ser Gly Ser Val His Ser
                245                 250                 255
Ala Ala Asp Ser Ser Trp Pro Thr Thr Pro Ser Gln Arg Gly Phe
                260                 265                 270
```

```
Leu Gln Arg Arg Ser Lys Phe Ser Arg Pro Glu Phe Ile Leu Leu
                275                 280                 285
Ala Gly Glu Ala Pro Met Thr Leu His Leu Pro Gly Ser Val Val
                290                 295                 300
Thr Lys Lys Gly Lys Ser Tyr Leu Ala Leu Arg Asp Leu Cys Val
                305                 310                 315
Val Leu Leu Asn Gly Gln His Leu Glu Val Lys Cys Asp Val Glu
                320                 325                 330
Ser Thr Val Gly Ala Val Phe Asn Ala Val Thr Ser Phe Ala Asn
                335                 340                 345
Leu Glu Glu Leu Thr Tyr Phe Gly Leu Ala Tyr Met Lys Ser Lys
                350                 355                 360
Glu Phe Phe Phe Leu Asp Ser Glu Thr Arg Leu Cys Lys Ile Ala
                365                 370                 375
Pro Glu Gly Trp Arg Glu Gln Pro Gln Lys Thr Ser Met Asn Thr
                380                 385                 390
Phe Thr Leu Phe Leu Arg Ile Lys Phe Phe Val Ser His Tyr Gly
                395                 400                 405
Leu Leu Gln His Ser Leu Thr Arg His Gln Phe Tyr Leu Gln Leu
                410                 415                 420
Arg Lys Asp Ile Leu Glu Glu Arg Leu Tyr Cys Asn Glu Glu Ile
                425                 430                 435
Leu Leu Gln Leu Gly Val Leu Ala Leu Gln Ala Glu Phe Gly Asn
                440                 445                 450
Tyr Pro Lys Glu Gln Val Glu Ser Lys Pro Tyr Phe His Val Glu
                455                 460                 465
Asp Tyr Ile Pro Ala Ser Leu Ile Glu Arg Met Thr Ala Leu Arg
                470                 475                 480
Val Gln Val Glu Val Ser Glu Met His Arg Leu Ser Ser Ala Leu
                485                 490                 495
Trp Gly Glu Asp Ala Glu Leu Lys Phe Leu Arg Val Thr Gln Gln
                500                 505                 510
Leu Pro Glu Tyr Gly Val Leu Val His Gln Val Phe Ser Glu Lys
                515                 520                 525
Arg Arg Pro Glu Glu Glu Met Ala Leu Gly Ile Cys Ala Lys Gly
                530                 535                 540
Val Ile Val Tyr Glu Val Lys Asn Asn Ser Arg Ile Ala Met Leu
                545                 550                 555
Arg Phe Gln Trp Arg Glu Thr Gly Lys Ile Ser Thr Tyr Gln Lys
                560                 565                 570
Lys Phe Thr Ile Thr Ser Ser Val Thr Gly Lys Lys His Thr Phe
                575                 580                 585
Val Thr Asp Ser Ala Lys Thr Ser Lys Tyr Leu Leu Asp Leu Cys
                590                 595                 600
Ser Ala Gln His Gly Phe Asn Ala Gln Met Gly Ser Gly Gln Pro
                605                 610                 615
Ser His Val Leu Phe Asp His Asp Lys Phe Val Gln Met Ala Asn
                620                 625                 630
Leu Ser Pro Ala His Gln Ala Arg Ser Lys Pro Leu Ile Trp Ile
                635                 640                 645
Gln Arg Leu Ser Cys Ser Glu Asn Glu Leu Phe Val Ser Arg Leu
                650                 655                 660
Gln Gly Ala Ala Gly Gly Leu Leu Ser Thr Ser Met Asp Asn Phe
                665                 670                 675
Asn Val Asp Gly Ser Lys Glu Ala Gly Ala Glu Gly Ile Gly Arg
                680                 685                 690
Ser Pro Cys Thr Gly Arg Glu Gln Leu Lys Ser Ala Cys Val Ile
                695                 700                 705
Gln Lys Pro Met Thr Trp Asp Ser Leu Ser Gly Pro Pro Val Gln
                710                 715                 720
Ser Met His Ala Gly Ser Lys Asn Asn Arg Arg Lys Ser Phe Ile
                725                 730                 735
Ala Glu Pro Gly Arg Glu Ile Val Arg Val Thr Leu Lys Arg Asp
```

```
                740                 745                 750
Pro His Arg Gly Phe Gly Phe Val Ile Asn Glu Gly Glu Tyr Ser
                755                 760                 765
Gly Gln Ala Asp Pro Gly Ile Phe Ile Ser Ser Ile Ile Pro Gly
                770                 775                 780
Gly Pro Ala Glu Lys Ala Lys Thr Ile Lys Pro Gly Gly Gln Ile
                785                 790                 795
Leu Ala Leu Asn His Ile Ser Leu Glu Gly Phe Thr Phe Asn Met
                800                 805                 810
Ala Val Arg Met Ile Gln Asn Ser Pro Asp Asn Ile Glu Leu Ile
                815                 820                 825
Ile Ser Gln Ser Lys Gly Val Gly Gly Asn Asn Pro Asp Glu Glu
                830                 835                 840
Lys Asn Ser Thr Ala Asn Ser Gly Val Ser Ser Thr Asp Ile Leu
                845                 850                 855
Ser Phe Gly Tyr Gln Gly Ser Leu Leu Ser His Thr Gln Asp Gln
                860                 865                 870
Asp Arg Asn Thr Glu Glu Leu Asp Met Ala Gly Val Gln Ser Leu
                875                 880                 885
Val Pro Arg Leu Arg His Gln Leu Ser Phe Leu Pro Leu Lys Gly
                890                 895                 900
Ala Gly Ser Ser Cys Pro Pro Ser Pro Pro Glu Ile Ser Ala Gly
                905                 910                 915
Glu Ile Tyr Phe Val Glu Leu Val Lys Glu Asp Gly Thr Leu Gly
                920                 925                 930
Phe Ser Val Thr Gly Gly Ile Asn Thr Ser Val Pro Tyr Gly Gly
                935                 940                 945
Ile Tyr Val Lys Ser Ile Val Pro Gly Gly Pro Ala Ala Lys Glu
                950                 955                 960
Gly Gln Ile Leu Gln Gly Asp Arg Leu Leu Gln Val Asp Gly Val
                965                 970                 975
Ile Leu Cys Gly Leu Thr His Lys Gln Ala Val Gln Cys Leu Lys
                980                 985                 990
Gly Pro Gly Gln Val Ala Arg Leu Val Leu Glu Arg Arg Val Pro
                995                1000                1005
Arg Ser Thr Gln Gln Cys Pro Ser Ala Asn Asp Ser Met Gly Asp
               1010                1015                1020
Glu Arg Thr Ala Val Ser Leu Val Thr Ala Leu Pro Gly Arg Pro
               1025                1030                1035
Ser Ser Cys Val Ser Val Thr Asp Gly Pro Lys Phe Glu Val Lys
               1040                1045                1050
Leu Lys Lys Asn Ala Asn Gly Leu Gly Phe Ser Phe Val Gln Met
               1055                1060                1065
Glu Lys Glu Ser Cys Ser His Leu Lys Ser Asp Leu Val Arg Ile
               1070                1075                1080
Lys Arg Leu Phe Pro Gly Gln Pro Ala Glu Glu Asn Gly Ala Ile
               1085                1090                1095
Ala Ala Gly Asp Ile Ile Leu Ala Val Asn Gly Arg Ser Thr Glu
               1100                1105                1110
Gly Leu Ile Phe Gln Glu Val Leu His Leu Leu Arg Gly Ala Pro
               1115                1120                1125
Gln Glu Val Thr Leu Leu Leu Cys Arg Pro Pro Pro Gly Ala Leu
               1130                1135                1140
Pro Glu Leu Glu Gln Glu Trp Gln Thr Pro Glu Leu Ser Ala Asp
               1145                1150                1155
Lys Glu Phe Thr Arg Ala Thr Cys Thr Asp Ser Cys Thr Ser Pro
               1160                1165                1170
Ile Leu Asp Gln Glu Asp Ser Trp Arg Asp Ser Ala Ser Pro Asp
               1175                1180                1185
Ala Gly Glu Gly Leu Gly Leu Arg Pro Glu Ser Ser Gln Lys Ala
               1190                1195                1200
Ile Arg Glu Ala Gln Trp Gly Gln Asn Arg Glu Arg Pro Trp Ala
               1205                1210                1215
```

```
Ser Ser Leu Thr His Ser Pro Glu Ser His Pro His Leu Cys Lys
                1220                1225                1230
Leu His Gln Glu Arg Asp Glu Ser Thr Leu Ala Thr Ser Leu Glu
                1235                1240                1245
Lys Asp Val Arg Gln Asn Cys Tyr Ser Val Cys Asp Ile Met Arg
                1250                1255                1260
Leu Gly Arg Tyr Ser Phe Ser Ser Pro Leu Thr Arg Leu Ser Thr
                1265                1270                1275
Asp Ile Phe


<210> 5
<211> 218
<212> PRT
<213> Homo sapiens

<220>
<221> misc_feature
<223> Incyte ID No: 6282188CD1

<400> 5
Met Leu Lys His Pro Val Leu Pro Ala Leu Cys Leu Ala Leu Val
 1               5                  10                  15
Ser Leu Phe Ala Asn Val Ser Val Gln Ala Asp Ala Ile Val Thr
                20                  25                  30
Ser Val Arg Ser Pro Glu Trp Ala Gln Pro Ile Asp Ala His Tyr
                35                  40                  45
Asn Leu His Gln Met Thr Pro Thr Leu Tyr Arg Ser Gly Leu Pro
                50                  55                  60
Asp Ser Arg Ala Leu Pro Leu Leu Glu Lys Leu Asn Val Gly Thr
                65                  70                  75
Val Ile Asn Phe Leu Pro Glu Ser Asp Asp Ser Trp Leu Ala Asp
                80                  85                  90
Ser Asp Ile Lys Gln Val Gln Leu Thr Tyr Arg Thr Asn His Val
                95                 100                 105
Asp Asp Ser Asp Val Leu Ala Ala Leu Arg Ala Ile Arg Gln Ala
                110                 115                 120
Glu Ala Asn Gly Ser Val Leu Met His Cys Lys His Gly Ser Asp
                125                 130                 135
Arg Thr Gly Leu Met Ala Ala Met Tyr Arg Val Val Ile Gln Gly
                140                 145                 150
Trp Ser Lys Glu Asp Ala Leu Asn Glu Met Thr Leu Gly Gly Phe
                155                 160                 165
Gly Ser Ser Asn Gly Phe Lys Asp Gly Val Arg Tyr Met Met Arg
                170                 175                 180
Ala Asp Ile Asp Lys Leu Arg Thr Ala Leu Ala Thr Gly Asp Cys
                185                 190                 195
Ser Thr Ser Ala Phe Ala Leu Cys Ser Met Lys Gln Trp Ile Ser
                200                 205                 210
Thr Thr Gly Ser Glu Gln Lys Glu
                215

<210> 6
<211> 420
<212> PRT
<213> Homo sapiens

<220>
<221> misc_feature
<223> Incyte ID No: 2182961CD1

<400> 6
Met Val Ala Ala Arg Glu Asn Glu Glu Glu Ala Lys Glu Glu Thr
```

```
1               5                   10                  15
Pro Asp Lys Leu Ile Arg Ser Cys Glu Pro Gly Ser Met Thr Val
                20                  25                  30
Thr Trp Thr Thr Trp Val Pro Thr Arg Ser Glu Val Gln Phe Gly
                35                  40                  45
Leu Gln Pro Ser Gly Pro Leu Pro Leu Arg Ala Gln Gly Thr Phe
                50                  55                  60
Val Pro Phe Val Asp Gly Gly Ile Leu Arg Arg Lys Leu Tyr Ile
                65                  70                  75
His Arg Val Thr Leu Arg Lys Leu Leu Pro Gly Val Gln Tyr Val
                80                  85                  90
Tyr Arg Cys Gly Ser Ala Gln Gly Trp Ser Arg Arg Phe Arg Phe
                95                  100                 105
Arg Ala Leu Lys Asn Gly Ala His Trp Ser Pro Arg Leu Ala Val
                110                 115                 120
Phe Gly Asp Leu Gly Ala Asp Asn Pro Lys Ala Val Pro Arg Leu
                125                 130                 135
Arg Arg Asp Thr Gln Gln Gly Met Tyr Asp Ala Val Leu His Val
                140                 145                 150
Gly Asp Phe Ala Tyr Asn Leu Asp Gln Asp Asn Ala Arg Val Gly
                155                 160                 165
Asp Arg Phe Met Arg Leu Ile Glu Pro Val Ala Ala Ser Leu Pro
                170                 175                 180
Tyr Met Thr Cys Pro Gly Asn His Glu Glu Arg Tyr Asn Phe Ser
                185                 190                 195
Asn Tyr Lys Ala Arg Phe Ser Met Pro Gly Asp Asn Glu Gly Leu
                200                 205                 210
Trp Tyr Ser Trp Asp Leu Gly Pro Ala His Ile Ile Ser Phe Ser
                215                 220                 225
Thr Glu Val Tyr Phe Phe Leu His Tyr Gly Arg His Leu Val Gln
                230                 235                 240
Arg Gln Phe Arg Trp Leu Glu Ser Asp Leu Gln Lys Ala Asn Lys
                245                 250                 255
Asn Arg Ala Ala Arg Pro Trp Ile Ile Thr Met Gly His Arg Pro
                260                 265                 270
Met Tyr Cys Ser Asn Ala Asp Leu Asp Asp Cys Thr Arg His Glu
                275                 280                 285
Ser Lys Val Arg Lys Gly Leu Gln Gly Lys Leu Tyr Gly Leu Glu
                290                 295                 300
Asp Leu Phe Tyr Lys Tyr Gly Val Asp Leu Gln Leu Trp Ala His
                305                 310                 315
Glu His Ser Tyr Glu Arg Leu Trp Pro Ile Tyr Asn Tyr Gln Val
                320                 325                 330
Phe Asn Gly Ser Arg Glu Met Pro Tyr Thr Asn Pro Arg Gly Pro
                335                 340                 345
Val His Ile Ile Thr Gly Ser Ala Gly Cys Glu Glu Arg Leu Thr
                350                 355                 360
Pro Phe Ala Val Phe Pro Arg Pro Trp Ser Ala Val Arg Val Lys
                365                 370                 375
Glu Tyr Gly Tyr Thr Arg Leu His Ile Leu Asn Gly Thr His Ile
                380                 385                 390
His Ile Gln Gln Val Ser Asp Asp Gln Asp Gly Lys Ile Val Asp
                395                 400                 405
Asp Val Trp Val Val Arg Pro Leu Phe Gly Arg Arg Met Tyr Leu
                410                 415                 420
```

```
<210> 7
<211> 986
<212> PRT
<213> Homo sapiens

<220>
```

```
<221> misc_feature
<223> Incyte ID No: 5119906CD1

<400> 7
Met Arg Phe Phe Leu Arg Glu Ala Gly Thr Val Ser Ala Gly Thr
 1               5                  10                  15
Ser Gln Cys Pro Arg Ser Ser Trp Glu Leu Cys Leu Leu Ser Cys
                20                  25                  30
Pro Leu Pro Ser Val Ser Cys Glu Met Arg Gly Leu Arg Leu Gln
                35                  40                  45
Ser Leu Ser Thr Leu Trp Thr Leu Ile Met Cys Val Val Pro Thr
                50                  55                  60
Arg Ala His Val Val Leu Ala Pro Ser Tyr Pro Asp Val Thr Phe
                65                  70                  75
Thr Ala Gly Ala Asp Phe Ser Pro Gln Ile Pro Phe Ser Leu Cys
                80                  85                  90
Phe Ile Leu Ser Gly Phe Ser Val Ser Thr Ala Gly Arg Met His
                95                 100                 105
Ile Phe Lys Pro Val Ser Val Gln Ala Met Trp Ser Ala Leu Gln
               110                 115                 120
Val Leu His Lys Ala Cys Glu Val Ala Arg Arg His Asn Tyr Phe
               125                 130                 135
Pro Gly Gly Val Ala Leu Ile Trp Ala Thr Tyr Tyr Glu Ser Cys
               140                 145                 150
Ile Ser Ser Glu Gln Ser Cys Ile Asn Glu Trp Asn Ala Met Gln
               155                 160                 165
Asp Leu Glu Ser Thr Arg Pro Asp Ser Pro Ala Leu Phe Val Asp
               170                 175                 180
Lys Pro Thr Glu Gly Glu Arg Thr Glu Arg Leu Ile Lys Ala Lys
               185                 190                 195
Leu Arg Ser Ile Met Met Ser Gln Asp Leu Glu Asn Val Thr Ser
               200                 205                 210
Lys Glu Ile Arg Asn Glu Leu Glu Lys Gln Met Asn Cys Asn Leu
               215                 220                 225
Lys Glu Leu Lys Glu Phe Ile Asp Asn Glu Met Leu Leu Ile Leu
               230                 235                 240
Gly Gln Met Asp Lys Pro Ser Leu Ile Phe Asp His Leu Tyr Leu
               245                 250                 255
Gly Ser Glu Trp Asn Ala Ser Asn Leu Glu Glu Leu Gln Gly Ser
               260                 265                 270
Gly Val Asp Tyr Ile Leu Asn Val Thr Arg Glu Ile Asp Asn Phe
               275                 280                 285
Phe Pro Gly Leu Phe Ala Tyr His Asn Ile Arg Val Tyr Asp Glu
               290                 295                 300
Glu Thr Thr Asp Leu Leu Ala His Trp Asn Glu Ala Tyr His Phe
               305                 310                 315
Ile Asn Lys Ala Lys Arg Asn His Ser Lys Cys Leu Val His Cys
               320                 325                 330
Lys Met Gly Val Ser Arg Ser Ala Ser Thr Val Ile Ala Tyr Ala
               335                 340                 345
Met Lys Glu Phe Gly Trp Pro Leu Glu Lys Ala Tyr Asn Tyr Val
               350                 355                 360
Lys Gln Lys Arg Ser Ile Thr Arg Pro Asn Ala Gly Phe Met Arg
               365                 370                 375
Gln Leu Ser Glu Tyr Glu Gly Ile Leu Asp Ala Ser Lys Gln Arg
               380                 385                 390
His Asn Lys Leu Trp Arg Gln Gln Thr Asp Ser Ser Leu Gln Gln
               395                 400                 405
Pro Val Asp Asp Pro Ala Gly Pro Gly Asp Phe Leu Pro Glu Thr
               410                 415                 420
Pro Asp Gly Thr Pro Glu Ser Gln Leu Pro Phe Leu Asp Asp Ala
               425                 430                 435
Ala Gln Pro Gly Leu Gly Pro Pro Leu Pro Cys Cys Phe Arg Arg
```

```
                440                 445                 450
Leu Ser Asp Pro Leu Leu Pro Ser Pro Glu Asp Glu Thr Gly Ser
                455                 460                 465
Leu Val His Leu Glu Asp Pro Glu Arg Glu Ala Leu Leu Glu Glu
                470                 475                 480
Ala Ala Pro Pro Ala Glu Val His Arg Pro Ala Arg Gln Pro Gln
                485                 490                 495
Gln Gly Ser Gly Leu Cys Glu Lys Asp Val Lys Lys Lys Leu Glu
                500                 505                 510
Phe Gly Ser Pro Lys Gly Arg Ser Gly Ser Leu Leu Gln Val Glu
                515                 520                 525
Glu Thr Glu Arg Glu Glu Gly Leu Gly Ala Gly Arg Trp Gly Gln
                530                 535                 540
Leu Pro Thr Gln Leu Asp Gln Asn Leu Leu Asn Ser Glu Asn Leu
                545                 550                 555
Asn Asn Asn Ser Lys Arg Ser Cys Pro Asn Gly Met Glu Asp Asp
                560                 565                 570
Ala Ile Phe Gly Ile Leu Asn Lys Val Lys Pro Ser Tyr Lys Ser
                575                 580                 585
Cys Ala Asp Cys Met Tyr Pro Thr Ala Ser Gly Ala Pro Glu Ala
                590                 595                 600
Ser Arg Glu Arg Cys Glu Asp Pro Asn Ala Pro Ala Ile Cys Thr
                605                 610                 615
Gln Pro Ala Phe Leu Pro His Ile Thr Ser Ser Pro Val Ala His
                620                 625                 630
Leu Ala Ser Arg Ser Arg Val Pro Glu Lys Pro Ala Ser Gly Pro
                635                 640                 645
Thr Glu Pro Pro Pro Phe Leu Pro Pro Ala Gly Ser Arg Arg Ala
                650                 655                 660
Asp Thr Ser Gly Pro Gly Ala Gly Ala Ala Leu Glu Pro Pro Ala
                665                 670                 675
Ser Leu Leu Glu Pro Ser Arg Glu Thr Pro Lys Val Leu Pro Lys
                680                 685                 690
Ser Leu Leu Leu Lys Asn Ser His Cys Asp Lys Asn Pro Pro Ser
                695                 700                 705
Thr Glu Val Val Ile Lys Glu Glu Ser Ser Pro Lys Lys Asp Met
                710                 715                 720
Lys Pro Ala Lys Asp Leu Arg Leu Leu Phe Ser Asn Glu Ser Glu
                725                 730                 735
Lys Pro Thr Thr Asn Ser Tyr Leu Met Gln His Gln Glu Ser Ile
                740                 745                 750
Ile Gln Leu Gln Lys Ala Gly Leu Val Arg Lys His Thr Lys Glu
                755                 760                 765
Leu Glu Arg Leu Lys Ser Val Pro Ala Asp Pro Ala Pro Pro Ser
                770                 775                 780
Arg Asp Gly Pro Ala Ser Arg Leu Glu Ala Ser Ile Pro Glu Glu
                785                 790                 795
Ser Gln Asp Pro Ala Ala Leu His Glu Leu Gly Pro Leu Val Met
                800                 805                 810
Pro Ser Gln Ala Gly Ser Asp Glu Lys Ser Glu Ala Ala Pro Ala
                815                 820                 825
Ser Leu Glu Gly Gly Ser Leu Lys Ser Pro Pro Pro Phe Phe Tyr
                830                 835                 840
Arg Leu Asp His Thr Ser Ser Phe Ser Lys Asp Phe Leu Lys Thr
                845                 850                 855
Ile Cys Tyr Thr Pro Thr Ser Ser Ser Met Ser Ser Asn Leu Thr
                860                 865                 870
Arg Ser Ser Ser Ser Asp Ser Ile His Ser Val Arg Gly Lys Pro
                875                 880                 885
Gly Leu Val Lys Gln Arg Thr Gln Glu Ile Glu Thr Arg Leu Arg
                890                 895                 900
Leu Ala Gly Leu Thr Val Ser Ser Pro Leu Lys Arg Ser His Ser
                905                 910                 915
```

```
Leu Ala Lys Leu Gly Ser Leu Thr Phe Ser Thr Glu Asp Leu Ser
                920                 925                 930
Ser Glu Ala Asp Pro Ser Thr Val Ala Asp Ser Gln Asp Thr Thr
                935                 940                 945
Leu Ser Glu Ser Ser Phe Leu His Glu Pro Gln Gly Thr Pro Arg
                950                 955                 960
Asp Pro Ala Ala Thr Ser Lys Pro Ser Gly Lys Pro Ala Pro Glu
                965                 970                 975
Asn Leu Lys Ser Pro Ser Trp Met Ser Lys Ser
                980                 985

<210> 8
<211> 399
<212> PRT
<213> Homo sapiens

<220>
<221> misc_feature
<223> Incyte ID No: 4022502CD1

<400> 8
Met Ala Glu Leu Leu Arg Ser Leu Gln Asp Ser Gln Leu Val Ala
 1               5                  10                  15
Arg Phe Gln Arg Arg Cys Gly Leu Phe Pro Ala Pro Asp Glu Gly
                20                  25                  30
Pro Arg Glu Asn Gly Ala Asp Pro Thr Glu Arg Ala Ala Arg Val
                35                  40                  45
Pro Gly Val Glu His Leu Pro Ala Ala Asn Gly Lys Gly Gly Glu
                50                  55                  60
Ala Pro Ala Asn Gly Leu Arg Arg Ala Ala Ala Pro Glu Ala Tyr
                65                  70                  75
Val Gln Lys Tyr Val Val Lys Asn Tyr Phe Tyr Tyr Tyr Leu Phe
                80                  85                  90
Gln Phe Ser Ala Ala Leu Gly Gln Glu Val Phe Tyr Ile Thr Phe
                95                  100                 105
Leu Pro Phe Thr His Trp Asn Ile Asp Pro Tyr Leu Ser Arg Arg
                110                 115                 120
Leu Ile Ile Ile Trp Val Leu Val Met Tyr Ile Gly Gln Val Ala
                125                 130                 135
Lys Asp Val Leu Lys Trp Pro Arg Pro Ser Ser Pro Pro Val Val
                140                 145                 150
Lys Leu Glu Lys Arg Leu Ile Ala Glu Tyr Gly Met Pro Ser Thr
                155                 160                 165
His Ala Met Ala Ala Thr Ala Ile Ala Phe Thr Leu Leu Ile Ser
                170                 175                 180
Thr Met Asp Arg Tyr Gln Tyr Pro Phe Val Leu Gly Leu Val Met
                185                 190                 195
Ala Val Val Phe Ser Thr Leu Val Cys Leu Ser Arg Leu Tyr Thr
                200                 205                 210
Gly Met His Thr Val Leu Asp Val Leu Gly Gly Val Leu Ile Thr
                215                 220                 225
Ala Leu Leu Ile Val Leu Thr Tyr Pro Ala Trp Thr Phe Ile Asp
                230                 235                 240
Cys Leu Asp Ser Ala Ser Pro Leu Phe Pro Val Cys Val Ile Val
                245                 250                 255
Val Pro Phe Phe Leu Cys Tyr Asn Tyr Pro Val Ser Asp Tyr Tyr
                260                 265                 270
Ser Pro Thr Arg Ala Asp Thr Thr Thr Ile Leu Ala Ala Gly Ala
                275                 280                 285
Gly Val Thr Ile Gly Phe Trp Ile Asn His Phe Phe Gln Leu Val
                290                 295                 300
Ser Lys Pro Ala Glu Ser Leu Pro Val Ile Gln Asn Ile Pro Pro
                305                 310                 315
```

```
Leu Thr Thr Tyr Met Leu Val Leu Gly Leu Thr Lys Phe Ala Val
                320                 325                 330
Gly Ile Val Leu Ile Leu Leu Val Arg Gln Leu Val Gln Asn Leu
                335                 340                 345
Ser Leu Gln Val Leu Tyr Ser Trp Phe Lys Val Val Thr Arg Asn
                350                 355                 360
Lys Glu Ala Arg Arg Arg Leu Glu Ile Glu Val Pro Tyr Lys Phe
                365                 370                 375
Val Thr Tyr Thr Ser Val Gly Ile Cys Ala Thr Thr Phe Val Pro
                380                 385                 390
Met Leu His Arg Phe Leu Gly Leu Pro
                395

<210> 9
<211> 387
<212> PRT
<213> Homo sapiens

<220>
<221> misc_feature
<223> Incyte ID No: 4084356CD1

<400> 9
Met Arg Ala Trp Ile Pro Gly Trp Val Gly Arg Pro His Gly Gly
 1               5                  10                  15
Ala Glu Ala Ser Gly Gly Leu Arg Phe Gly Ala Ser Ala Ala Gln
                20                  25                  30
Gly Trp Arg Ala Arg Met Glu Asp Ala His Cys Thr Trp Leu Ser
                35                  40                  45
Leu Pro Gly Leu Pro Pro Gly Trp Ala Leu Phe Ala Val Leu Asp
                50                  55                  60
Gly His Gly Gly Ala Arg Ala Ala Arg Phe Gly Ala Arg His Leu
                65                  70                  75
Pro Gly His Val Leu Gln Glu Leu Gly Pro Glu Pro Ser Glu Pro
                80                  85                  90
Glu Gly Val Arg Glu Ala Leu Arg Arg Ala Phe Leu Ser Ala Asp
                95                  100                 105
Glu Arg Leu Arg Ser Leu Trp Pro Arg Val Glu Thr Gly Gly Phe
                110                 115                 120
Thr Ala Val Val Leu Leu Val Ser Pro Arg Phe Leu Tyr Leu Ala
                125                 130                 135
His Cys Gly Asp Ser Arg Ala Val Leu Ser Arg Ala Gly Ala Val
                140                 145                 150
Ala Phe Ser Thr Glu Asp His Arg Pro Leu Arg Pro Arg Glu Arg
                155                 160                 165
Glu Arg Ile His Ala Ala Gly Gly Thr Ile Arg Arg Arg Arg Val
                170                 175                 180
Glu Gly Ser Leu Ala Val Ser Arg Ala Leu Gly Asp Phe Thr Tyr
                185                 190                 195
Lys Glu Ala Pro Gly Arg Pro Pro Glu Leu Gln Leu Val Ser Ala
                200                 205                 210
Glu Pro Glu Val Ala Ala Leu Ala Arg Gln Ala Glu Asp Glu Phe
                215                 220                 225
Met Leu Leu Ala Ser Asp Gly Val Trp Asp Thr Val Ser Gly Ala
                230                 235                 240
Ala Leu Ala Gly Leu Val Ala Ser Arg Leu Arg Leu Gly Leu Ala
                245                 250                 255
Pro Glu Leu Leu Cys Ala Gln Leu Leu Asp Thr Cys Leu Cys Lys
                260                 265                 270
Gly Ser Leu Asp Asn Met Thr Cys Ile Leu Val Cys Phe Pro Gly
                275                 280                 285
Ala Pro Arg Pro Ser Glu Glu Ala Ile Arg Arg Glu Leu Ala Leu
                290                 295                 300
```

Asp Ala Ala Leu Gly Cys Arg Ile Ala Glu Leu Cys Ala Ser Ala
305 310 315
Gln Lys Pro Pro Ser Leu Asn Thr Val Phe Arg Thr Leu Ala Ser
320 325 330
Glu Asp Ile Pro Asp Leu Pro Pro Gly Gly Gly Leu Asp Cys Lys
335 340 345
Ala Thr Val Ile Ala Glu Val Tyr Ser Gln Ile Cys Gln Val Ser
350 355 360
Glu Glu Cys Gly Glu Lys Gly Gln Asp Gly Ala Gly Lys Ser Asn
365 370 375
Pro Thr His Leu Gly Ser Ala Leu Asp Met Glu Ala
380 385

<210> 10
<211> 447
<212> PRT
<213> Homo sapiens

<220>
<221> misc_feature
<223> Incyte ID No: 1740204CD1

<400> 10
Met Gly Glu Asp Thr Asp Thr Arg Lys Ile Asn His Ser Phe Leu
1 5 10 15
Arg Asp His Ser Tyr Val Thr Glu Ala Asp Ile Phe Ser Thr Val
20 25 30
Glu Phe Asn His Thr Gly Glu Leu Leu Ala Thr Gly Asp Lys Gly
35 40 45
Gly Arg Val Val Ile Phe Gln Arg Glu Pro Glu Ser Lys Asn Ala
50 55 60
Pro His Ser Gln Gly Glu Tyr Asp Val Tyr Ser Thr Phe Gln Ser
65 70 75
His Glu Pro Glu Phe Asp Tyr Leu Lys Ser Leu Glu Ile Glu Glu
80 85 90
Lys Ile Asn Lys Ile Lys Trp Leu Pro Gln Gln Asn Ala Ala His
95 100 105
Ser Leu Leu Ser Thr Asn Asp Lys Thr Ile Lys Leu Trp Lys Ile
110 115 120
Thr Glu Arg Asp Lys Arg Pro Glu Gly Tyr Asn Leu Lys Asp Glu
125 130 135
Glu Gly Lys Leu Lys Asp Leu Ser Thr Val Thr Ser Leu Gln Val
140 145 150
Pro Val Leu Lys Pro Met Asp Leu Met Val Glu Val Ser Pro Arg
155 160 165
Arg Ile Phe Ala Asn Gly His Thr Tyr His Ile Asn Ser Ile Ser
170 175 180
Val Asn Ser Asp Cys Glu Thr Tyr Met Ser Ala Asp Asp Leu Arg
185 190 195
Ile Asn Leu Trp His Leu Ala Ile Thr Asp Arg Ser Phe Asn Ile
200 205 210
Val Asp Ile Lys Pro Ala Asn Met Glu Asp Leu Thr Glu Val Ile
215 220 225
Thr Ala Ser Glu Phe His Pro His His Cys Asn Leu Phe Val Tyr
230 235 240
Ser Ser Ser Lys Gly Ser Leu Arg Leu Cys Asp Met Arg Ala Ala
245 250 255
Ala Leu Cys Asp Lys His Ser Lys Leu Phe Glu Glu Pro Glu Asp
260 265 270
Pro Ser Asn Arg Ser Phe Phe Ser Glu Ile Ile Ser Ser Val Ser
275 280 285
Asp Val Lys Phe Ser His Ser Gly Arg Tyr Met Leu Thr Arg Asp
290 295 300

Tyr Leu Thr Val Lys Val Trp Asp Leu Asn Met Glu Ala Arg Pro
305 310 315
Ile Glu Thr Tyr Gln Val His Asp Tyr Leu Arg Ser Lys Leu Cys
320 325 330
Ser Leu Tyr Glu Asn Asp Cys Ile Phe Asp Lys Phe Glu Cys Ala
335 340 345
Trp Asn Gly Ser Asp Ser Val Ile Met Thr Gly Ala Tyr Asn Asn
350 355 360
Phe Phe Arg Met Phe Asp Arg Asn Thr Lys Arg Asp Val Thr Leu
365 370 375
Glu Ala Ser Arg Glu Ser Ser Lys Pro Arg Ala Val Leu Lys Pro
380 385 390
Arg Arg Val Cys Val Gly Gly Lys Arg Arg Arg Asp Asp Ile Ser
395 400 405
Val Asp Ser Leu Asp Phe Thr Lys Lys Ile Leu His Thr Ala Trp
410 415 420
His Pro Ala Glu Asn Ile Ile Ala Ile Ala Ala Thr Asn Asn Leu
425 430 435
Tyr Ile Phe Gln Asp Lys Val Asn Ser Asp Met His
440 445

<210> 11
<211> 572
<212> PRT
<213> Homo sapiens

<220>
<221> misc_feature
<223> Incyte ID No: 7483804CD1

<400> 11
Met Asn Tyr Glu Gly Ala Arg Ser Glu Arg Glu Asn His Ala Ala
1 5 10 15
Asp Asp Ser Glu Gly Gly Ala Leu Asp Met Cys Cys Ser Glu Arg
20 25 30
Leu Pro Gly Leu Pro Gln Pro Ile Val Met Glu Ala Leu Asp Glu
35 40 45
Ala Glu Gly Leu Gln Asp Ser Gln Arg Glu Met Pro Pro Pro Pro
50 55 60
Pro Pro Ser Pro Pro Ser Asp Pro Ala Gln Lys Pro Pro Pro Arg
65 70 75
Gly Ala Gly Ser His Ser Leu Thr Val Arg Ser Ser Leu Cys Leu
80 85 90
Phe Ala Ala Ser Gln Phe Leu Leu Ala Cys Gly Val Leu Trp Phe
95 100 105
Ser Gly Tyr Gly His Ile Trp Ser Gln Asn Ala Thr Asn Leu Val
110 115 120
Ser Ser Leu Leu Thr Leu Leu Lys Gln Leu Glu Pro Thr Ala Trp
125 130 135
Leu Asp Ser Gly Thr Trp Gly Val Pro Ser Leu Leu Leu Val Phe
140 145 150
Leu Ser Val Gly Leu Val Leu Val Thr Thr Leu Val Trp His Leu
155 160 165
Leu Arg Thr Pro Pro Glu Pro Pro Thr Pro Leu Pro Pro Glu Asp
170 175 180
Arg Arg Gln Ser Val Ser Arg Gln Pro Ser Phe Thr Tyr Ser Glu
185 190 195
Trp Met Glu Glu Lys Ile Glu Asp Asp Phe Leu Asp Leu Asp Pro
200 205 210
Val Pro Glu Thr Pro Val Phe Asp Cys Val Met Asp Ile Lys Pro
215 220 225
Glu Ala Asp Pro Thr Ser Leu Thr Val Lys Ser Met Gly Leu Gln
230 235 240

```
Glu Arg Arg Gly Ser Asn Val Ser Leu Thr Leu Asp Met Cys Thr
                245                 250                 255
Pro Gly Cys Asn Glu Glu Gly Phe Gly Tyr Leu Met Ser Pro Arg
                260                 265                 270
Glu Glu Ser Ala Arg Glu Tyr Leu Leu Ser Ala Ser Arg Val Leu
                275                 280                 285
Gln Ala Glu Glu Leu His Glu Lys Ala Leu Asp Pro Phe Leu Leu
                290                 295                 300
Gln Ala Glu Phe Phe Glu Ile Pro Met Asn Phe Val Asp Pro Lys
                305                 310                 315
Glu Tyr Asp Ile Pro Gly Leu Val Arg Lys Asn Arg Tyr Lys Thr
                320                 325                 330
Ile Leu Pro Asn Pro His Ser Arg Val Cys Leu Thr Ser Pro Asp
                335                 340                 345
Pro Asp Asp Pro Leu Ser Ser Tyr Ile Asn Ala Asn Tyr Ile Arg
                350                 355                 360
Pro Gly Leu Gly Trp Pro Gln Gly Tyr Gly Gly Glu Glu Lys Val
                365                 370                 375
Tyr Ile Ala Thr Gln Gly Pro Ile Val Ser Thr Val Ala Asp Phe
                380                 385                 390
Trp Arg Met Val Trp Gln Glu His Thr Pro Ile Ile Val Met Ile
                395                 400                 405
Thr Asn Ile Glu Glu Met Asn Glu Lys Cys Thr Glu Tyr Trp Pro
                410                 415                 420
Glu Glu Gln Val Ala Tyr Asp Gly Val Glu Ile Thr Val Gln Lys
                425                 430                 435
Val Ile His Thr Glu Asp Tyr Arg Leu Arg Leu Ile Ser Leu Lys
                440                 445                 450
Ser Gly Thr Glu Glu Arg Gly Leu Lys His Tyr Trp Phe Thr Ser
                455                 460                 465
Trp Pro Asp Gln Lys Thr Pro Asp Arg Ala Pro Pro Leu Leu His
                470                 475                 480
Leu Val Arg Glu Val Glu Glu Ala Ala Gln Gln Glu Gly Pro His
                485                 490                 495
Cys Ala Pro Ile Ile Val His Cys Ser Ala Gly Ile Gly Arg Thr
                500                 505                 510
Gly Cys Phe Ile Ala Thr Ser Ile Cys Cys Gln Gln Leu Arg Gln
                515                 520                 525
Glu Gly Val Val Asp Ile Leu Lys Thr Thr Cys Gln Leu Arg Gln
                530                 535                 540
Asp Arg Gly Gly Met Ile Gln Thr Cys Glu Gln Tyr Gln Phe Val
                545                 550                 555
His His Val Met Ser Leu Tyr Glu Lys Gln Leu Ser His Gln Ser
                560                 565                 570
Pro Glu


<210> 12
<211> 1510
<212> PRT
<213> Homo sapiens

<220>
<221> misc_feature
<223> Incyte ID No: 7483934CD1

<400> 12
Met Ala Leu Ser Lys Gly Leu Arg Leu Leu Gly Arg Leu Gly Ala
 1               5                  10                  15
Glu Gly Asp Cys Ser Val Leu Leu Glu Ala Arg Gly Arg Asp Asp
                20                  25                  30
Cys Leu Leu Phe Glu Ala Gly Thr Val Ala Thr Leu Asp Asp Cys
                35                  40                  45
```

```
Leu Leu Phe Glu Ala Gly Thr Val Ala Thr Leu Ala Pro Glu Glu
                 50                  55                  60
Lys Glu Val Ile Lys Gly Gln Tyr Gly Lys Leu Thr Asp Ala Tyr
                 65                  70                  75
Gly Cys Leu Gly Glu Leu Arg Leu Lys Ser Gly Gly Thr Ser Leu
                 80                  85                  90
Ser Phe Leu Val Leu Val Thr Gly Cys Thr Ser Val Gly Arg Ile
                 95                 100                 105
Pro Asp Ala Glu Ile Tyr Lys Ile Thr Ala Thr Asp Phe Tyr Pro
                110                 115                 120
Leu Gln Glu Glu Ala Lys Glu Glu Glu Arg Leu Ile Ala Leu Lys
                125                 130                 135
Lys Ile Leu Ser Ser Gly Val Phe Tyr Phe Ser Trp Pro Asn Asp
                140                 145                 150
Gly Ser Arg Phe Asp Leu Thr Val Arg Thr Gln Lys Gln Gly Asp
                155                 160                 165
Asp Ser Ser Glu Trp Gly Asn Ser Phe Phe Trp Asn Gln Leu Leu
                170                 175                 180
His Val Pro Leu Arg Gln His Gln Val Ser Cys Cys Asp Trp Leu
                185                 190                 195
Leu Lys Ile Ile Cys Gly Val Val Thr Ile Arg Thr Val Tyr Ala
                200                 205                 210
Ser His Lys Gln Ala Lys Ala Cys Leu Val Ser Arg Val Ser Cys
                215                 220                 225
Glu Arg Thr Gly Thr Arg Phe His Thr Arg Gly Val Asn Asp Asp
                230                 235                 240
Gly His Val Ser Asn Phe Val Glu Thr Glu Gln Met Ile Tyr Met
                245                 250                 255
Asp Asp Gly Val Ser Ser Phe Val Gln Ile Arg Gly Ser Val Pro
                260                 265                 270
Leu Phe Trp Glu Gln Pro Gly Leu Gln Val Gly Ser His His Leu
                275                 280                 285
Arg Leu His Lys Gly Leu Glu Ala Asn Ala Pro Ala Phe Asp Arg
                290                 295                 300
His Met Val Leu Leu Lys Glu Gln Tyr Gly Gln Gln Val Val Val
                305                 310                 315
Asn Leu Leu Gly Ser Arg Gly Gly Glu Glu Val Leu Asn Arg Ala
                320                 325                 330
Phe Lys Lys Leu Leu Trp Ala Ser Cys His Ala Gly Asp Thr Pro
                335                 340                 345
Met Ile Asn Phe Asp Phe His Gln Phe Ala Lys Gly Gly Lys Leu
                350                 355                 360
Glu Lys Leu Glu Thr Leu Leu Arg Pro Gln Leu Lys Leu His Trp
                365                 370                 375
Glu Asp Phe Asp Val Phe Thr Lys Gly Glu Asn Val Ser Pro Arg
                380                 385                 390
Phe Gln Lys Gly Thr Leu Arg Met Asn Cys Leu Asp Cys Leu Asp
                395                 400                 405
Arg Thr Asn Thr Val Gln Ser Phe Ile Ala Leu Glu Val Leu His
                410                 415                 420
Leu Gln Leu Lys Thr Leu Gly Leu Ser Ser Lys Pro Ile Val Asp
                425                 430                 435
Arg Phe Val Glu Ser Phe Lys Ala Met Trp Ser Leu Asn Gly His
                440                 445                 450
Ser Leu Ser Lys Val Phe Thr Gly Ser Arg Ala Leu Glu Gly Lys
                455                 460                 465
Ala Lys Val Gly Lys Leu Lys Asp Gly Ala Arg Ser Met Ser Arg
                470                 475                 480
Thr Ile Gln Ser Asn Phe Phe Asp Gly Val Lys Gln Glu Ala Ile
                485                 490                 495
Lys Leu Leu Leu Val Gly Asp Val Tyr Gly Glu Glu Val Ala Asp
                500                 505                 510
Lys Gly Gly Met Leu Leu Asp Ser Thr Ala Leu Leu Val Thr Pro
```

```
                515                 520                 525
Arg Ile Leu Lys Ala Met Thr Glu Arg Gln Ser Glu Phe Thr Asn
                530                 535                 540
Phe Lys Arg Ile Arg Ile Ala Met Gly Thr Trp Asn Val Asn Gly
                545                 550                 555
Gly Lys Gln Phe Arg Ser Asn Val Leu Arg Thr Ala Glu Leu Thr
                560                 565                 570
Asp Trp Leu Leu Asp Ser Pro Gln Leu Ser Gly Ala Thr Asp Ser
                575                 580                 585
Gln Asp Asp Ser Ser Pro Ala Asp Ile Phe Ala Val Gly Phe Glu
                590                 595                 600
Glu Met Val Glu Leu Ser Ala Gly Asn Ile Val Asn Ala Ser Thr
                605                 610                 615
Thr Asn Lys Lys Met Trp Gly Glu Gln Leu Gln Lys Ala Ile Ser
                620                 625                 630
Arg Ser His Arg Tyr Ile Leu Leu Thr Ser Ala Gln Leu Val Gly
                635                 640                 645
Val Cys Leu Tyr Ile Phe Val Arg Pro Tyr His Val Pro Phe Ile
                650                 655                 660
Arg Asp Val Ala Ile Asp Thr Val Lys Thr Gly Met Gly Gly Lys
                665                 670                 675
Ala Gly Asn Lys Gly Ala Val Gly Ile Arg Phe Gln Phe His Ser
                680                 685                 690
Thr Ser Phe Cys Phe Ile Cys Ser His Leu Thr Ala Gly Gln Ser
                695                 700                 705
Gln Val Lys Glu Arg Asn Glu Asp Tyr Lys Glu Ile Thr Gln Lys
                710                 715                 720
Leu Cys Phe Pro Met Gly Arg Asn Val Phe Ser His Asp Tyr Val
                725                 730                 735
Phe Trp Cys Gly Asp Phe Asn Tyr Arg Ile Asp Leu Thr Tyr Glu
                740                 745                 750
Glu Val Phe Tyr Phe Val Lys Arg Gln Asp Trp Lys Lys Leu Leu
                755                 760                 765
Glu Phe Asp Gln Leu Gln Leu Gln Lys Ser Ser Gly Lys Ile Phe
                770                 775                 780
Lys Asp Phe His Glu Gly Ala Ile Asn Phe Gly Pro Thr Tyr Lys
                785                 790                 795
Tyr Asp Val Gly Ser Ala Ala Tyr Asp Thr Ser Asp Lys Cys Arg
                800                 805                 810
Thr Pro Ala Trp Thr Asp Arg Val Leu Trp Trp Arg Lys Lys His
                815                 820                 825
Pro Phe Asp Lys Thr Ala Gly Glu Leu Asn Leu Leu Asp Ser Asp
                830                 835                 840
Leu Asp Val Asp Thr Lys Val Arg His Thr Trp Ser Pro Gly Ala
                845                 850                 855
Leu Gln Tyr Tyr Gly Arg Ala Glu Leu Gln Ala Ser Asp His Arg
                860                 865                 870
Pro Val Leu Ala Ile Val Glu Val Glu Val Gln Glu Val Asp Val
                875                 880                 885
Gly Ala Arg Glu Arg Val Phe Gln Glu Val Ser Ser Phe Gln Gly
                890                 895                 900
Pro Leu Asp Ala Thr Val Val Val Asn Leu Gln Ser Pro Thr Leu
                905                 910                 915
Glu Glu Lys Asn Glu Phe Pro Glu Asp Leu Arg Thr Glu Leu Met
                920                 925                 930
Gln Thr Leu Gly Ser Tyr Gly Thr Ile Val Leu Val Arg Ile Asn
                935                 940                 945
Gln Gly Gln Met Leu Val Thr Phe Ala Asp Ser His Ser Ala Leu
                950                 955                 960
Ser Val Leu Asp Val Asp Gly Met Lys Val Lys Gly Arg Ala Val
                965                 970                 975
Lys Ile Arg Pro Lys Thr Lys Asp Trp Leu Lys Gly Leu Arg Glu
                980                 985                 990
```

```
Glu Ile Ile Arg Lys Arg Asp Ser Met Ala Pro Val Ser Pro Thr
                995                 1000                1005
Ala Asn Ser Cys Leu Leu Glu Glu Asn Phe Asp Phe Thr Ser Leu
               1010                1015                1020
Asp Tyr Glu Ser Glu Gly Asp Ile Leu Glu Asp Asp Glu Asp Tyr
               1025                1030                1035
Leu Val Asp Glu Phe Asn Gln Pro Gly Val Ser Asp Ser Glu Leu
               1040                1045                1050
Gly Gly Asp Asp Leu Ser Asp Val Pro Gly Pro Thr Ala Leu Ala
               1055                1060                1065
Pro Pro Ser Lys Ser Pro Ala Leu Thr Lys Lys Lys Gln His Pro
               1070                1075                1080
Thr Tyr Lys Asp Asp Ala Asp Leu Val Glu Leu Lys Arg Glu Leu
               1085                1090                1095
Glu Ala Val Gly Glu Phe Arg His Arg Ser Pro Ser Arg Ser Leu
               1100                1105                1110
Ser Val Pro Asn Arg Pro Arg Pro Pro Gln Pro Pro Gln Arg Pro
               1115                1120                1125
Pro Pro Pro Thr Gly Leu Met Val Lys Lys Ser Ala Ser Asp Ala
               1130                1135                1140
Ser Ile Ser Ser Gly Thr His Gly Gln Tyr Ser Ile Leu Gln Thr
               1145                1150                1155
Ala Arg Leu Leu Pro Gly Ala Pro Gln Gln Pro Pro Lys Ala Arg
               1160                1165                1170
Thr Gly Ile Ser Lys Pro Tyr Asn Val Lys Gln Ile Lys Thr Thr
               1175                1180                1185
Asn Ala Gln Glu Ala Glu Ala Ala Ile Arg Cys Leu Leu Glu Ala
               1190                1195                1200
Arg Gly Gly Ala Ser Glu Glu Ala Leu Ser Ala Val Ala Pro Arg
               1205                1210                1215
Asp Leu Glu Ala Ser Ser Glu Pro Glu Pro Thr Pro Gly Ala Ala
               1220                1225                1230
Lys Pro Glu Thr Pro Gln Ala Pro Pro Leu Leu Pro Arg Arg Pro
               1235                1240                1245
Pro Pro Arg Val Pro Ala Ile Lys Lys Pro Thr Leu Arg Arg Thr
               1250                1255                1260
Gly Lys Pro Leu Ser Pro Glu Glu Gln Phe Glu Gln Gln Thr Val
               1265                1270                1275
His Phe Thr Ile Gly Pro Pro Glu Thr Ser Val Glu Ala Pro Pro
               1280                1285                1290
Val Val Thr Ala Pro Arg Val Pro Pro Val Pro Lys Pro Arg Thr
               1295                1300                1305
Phe Gln Pro Gly Lys Ala Ala Glu Arg Pro Ser His Arg Lys Pro
               1310                1315                1320
Ala Ser Asp Glu Ala Pro Pro Gly Ala Gly Ala Ser Val Pro Pro
               1325                1330                1335
Pro Leu Glu Ala Pro Pro Leu Val Pro Lys Val Pro Pro Arg Arg
               1340                1345                1350
Lys Lys Ser Ala Pro Ala Ala Phe His Leu Gln Val Leu Gln Ser
               1355                1360                1365
Asn Ser Gln Leu Leu Gln Gly Leu Thr Tyr Asn Ser Ser Asp Ser
               1370                1375                1380
Pro Ser Gly His Pro Pro Ala Ala Gly Thr Val Phe Pro Gln Gly
               1385                1390                1395
Asp Phe Leu Ser Thr Ser Ser Ala Thr Ser Pro Asp Ser Asp Gly
               1400                1405                1410
Thr Lys Ala Met Lys Pro Glu Ala Ala Pro Leu Leu Gly Asp Tyr
               1415                1420                1425
Gln Asp Pro Phe Trp Asn Leu Leu His His Pro Lys Leu Leu Asn
               1430                1435                1440
Asn Thr Trp Leu Ser Lys Ser Ser Asp Pro Leu Asp Ser Gly Thr
               1445                1450                1455
Arg Ser Pro Lys Arg Asp Pro Ile Asp Pro Val Ser Ala Gly Ala
```

```
                1460                1465                1470
Ser Ala Ala Lys Ala Glu Leu Pro Pro Asp His Gly His Lys Thr
                1475                1480                1485
Leu Gly His Trp Val Thr Ile Ser Asp Gln Glu Lys Arg Thr Ala
                1490                1495                1500
Leu Gln Val Phe Asp Pro Leu Ala Lys Thr
                1505                1510

<210> 13
<211> 1600
<212> DNA
<213> Homo sapiens

<220>
<221> misc_feature
<223> Incyte ID No: 3272350CB1

<400> 13
atggctggag agaatggcca ggagggagtg ggtatctgca ggttgggagt ccagccggag 60
gtggagccca gttcccagga cgtgcgccag gcgctgggcc ggcccgtgct cctgcgctgc 120
tcgctgctgc gaggcagccc ccagcgcatc gcctcggctg tgtggcgttt caaagggcag 180
ctgctgccgc cgccgcctgt tgttcccgcc gccgccgagg cgccggatca cgcggagctg 240
cgcctcgacg ccgtaactcg cgacagcagc ggcagctacg agtgcagcgt ctccaacgat 300
gtgggctcgg ctgcctgcct cttccaggtc tccgccaaag cctacagccc ggagttttac 360
ttcgacaccc ccaaccccac ccgcagccac aagctgtcca agaactactc ctacgtgctg 420
cagtggactc agagggagcc cgacgctgtc gaccctgtgc tcaactacag actcagcatc 480
cgccagttga accagcacaa tgcggtggtc aaggccatcc cggtccggcg tgtggagaag 540
gggcagctgc tggagtacat cctgaccgat ctccgtgtgc cccacagcta tgaggtccgc 600
ctcacaccct ataccacctt cggggctggt gacatggcct cccgcatcat ccactacaca 660
gagcccatca actctccgaa cctttcagac aacacctgcc actttgagga tgagaagatc 720
tgtggctata cccaggacct gacagacaac tttgactgga cgcggcagaa tgccctcacc 780
cagaacccca aacgctcccc caacactggt ccccccaccg acataagtgg cacccctgag 840
ggctactaca tgttcatcga gacatcgagg cctcgggagc tgggggaccg tgcaaggtta 900
gtgagtcccc tctacaatgc cagcgccaag ttctactgtg tctccttctt ctaccacatg 960
tacgggaaac acatcggctc cctcaacctc ctggtgcggt cccggaacaa aggggctctg 1020
gacacgcacg cctggtctct cagtggcaat aagggcaatg tgtggcagca ggcccatgtg 1080
cccatcagcc ccagtgggcc cttccagatt atttttgagg gggttcgagg cccgggctac 1140
ctgggggata ttgccataga tgacgtcaca ctgaagaagg gggagtgtcc ccggaagcag 1200
acggatccca ataaagtggt ggtgatgccg ggcagtggag ccccctgcca gtccagccca 1260
cagctgtggg ggcccatggc catcttcctc ttggcgttgc agagatgatg agagctgtgt 1320
ggccaccccc ccaaccttgc ccccggcaca ccaaagtgtc cacattgtac caaagactga 1380
cccccgccag ctggggtgcc caggggcagg gccggcccgc cagggagggg gcctgcattg 1440
gctgcaagga tgagcagaga acaaggacag aggccaggca ctgaggccct ggagacagct 1500
gttccacttg cacacacgca cacactcatg ctcacacaca cagagatata ttaaagcaca 1560
agtttctatc tgaaaaaaaa aaaaaaaaaa aaaaaaaaaa                       1600

<210> 14
<211> 781
<212> DNA
<213> Homo sapiens

<220>
<221> misc_feature
<223> Incyte ID No: 7481507CB1

<400> 14
aggcgcgccg tgaggaagaa ggggagccat ccgatatccc gaccctgaaa gactacaccg 60
cccgcctggt ggatcagaaa tggctgcgac tcgcggcgag gagaaaatct gcatgagcat 120
gtatcaacgc attaatggcg ctgactggcg caatattttc gtcgtcggcg atctgcatgg 180
gtgctacacg ctgctgatga atgaactcga aaaggtttcg ttcgaccctg cgtgtgattt 240
gctgatttcg gttggagacc ttgttgaccg cggcgcggaa aacgtcgagt gcctggagct 300
gattactatg ccttggttcc gggctgtgcg aggtaaccat gagcagatga tgattgatgg 360
gctatcggag tatggaaacg ttaaccactg gctggaaaac ggcggcgtgt ggttcttcag 420
```

tcttgattat gaaaaagagg tgctggctaa ggctctggtt cataaatcgg ccagcctgcc 480
attcgtcatc gagctggtta ccgctgaacg taaaatcgtt atctgccacg ctgactaccc 540
gcataacgaa tatgcgttcg acaagccggt cccgaaagac atggtcatct ggaatcgtga 600
acgggttagc gacgctcagg acggcattgt ctcgccgata gctggtgctg atctgtttat 660
cttcggccac acccctgcgc gccagcccct gaagtatgcc aaccagatgt acatcgatac 720
tggtgccgtg ttctgcggaa acctcacgct ggtacaggtt caaggtggtg cccatgcgta 780
a 781

<210> 15
<211> 1724
<212> DNA
<213> Homo sapiens

<220>
<221> misc_feature
<223> Incyte ID No: 2285140CB1

<400> 15
gtcgggaccc gtagtagagg cgggttgcgg gcggcggcgg cggcggcggc ggtggtggtt 60
gtggcgaggc tgtgcggcag ggcgcacggg acctgtcctg cagcggctct ctcaggccgt 120
gggtcgtcgc tgcagctgcc gggaaagaag gaaacgacga ctccgggggc gaacttggca 180
cacagggagg aagggaaagg atgcatggtc atggaggcta tgattctgat tttagtgatg 240
atgaacactg tggagaatcc agcaaaagga aaaaaaggac agttgaagac gacttactgc 300
tccaaaaacc atttcagaaa gaaaaacatg gaaaggtggc ccataaacaa gttgcagcag 360
aattgctgga tagggaagaa gcaagaaata gaaggtttca tctcatagct atggatgctt 420
atcaaagaca tagaaagttc gtaaatgact atattttata ctatggtggc aaaaaagaag 480
acttcaagcg tttgggggaa aatgacaaga cagacttgga tgttatacga gaaaatcata 540
gattcctatg gaatgaggag gacgaaatgg acatgacttg ggagaagaga cttgctaaga 600
aatactatga taaattattt aaggaatact gcatagcaga tctcagtaaa tataaagaaa 660
ataagtttgg atttaggtgg cgagtagaaa aagaagtaat ttcaggaaaa ggtcaatttt 720
tctgtggaaa taaatattgt gataaaaaag aaggcttaaa gagttgggaa gttaattttg 780
gttatattga gcatggtgag aagagaaatg cacttgttaa attaaggtta tgccaagaat 840
gttccattaa attaaatttc catcacagga gaaaagaaat caagtcaaaa aaaagaaaag 900
ataaaaccaa aaaagactgt gaagagtcat cacataaaaa atccagatta tcttctgcag 960
aagaggcctc caagaaaaaa gataaaggac attcatcttc aaagaaatct gaagattctc 1020
tacttagaaa ctctgatgag gaagaaagtg cttcagaatc tgaactttgg aagggtccac 1080
taccagagac agatgaaaaa tcacaggaag aagaatttga tgagtatttt caggatttgt 1140
ttctatgaga cgagagagag aagcctccgc tccttaatgt gaaacttcat gaagttttaa 1200
acttcatgca atttgaaatt ccatataagt ttttatctgc aagttacagc ttgtgtggtt 1260
tgtctttgga aataaaaatc caggttctct cagaatgtca gaggctttgg aagttcatta 1320
gttcaattaa agactttcct gtcctttaaa tatcttttca attgcttatc tacaattctg 1380
gtttatttgt agctcctaga ggatagagct ggacagattc cattgttcct acattttgta 1440
ggtttttttt cactgccttc attatggatc ttctcttgcc ttcattattt tattttaata 1500
attcttcttt ttctcttttt tagagccacc aataccggaa ttggttggct ttcatttttt 1560
tcctttgtgg aaacggagtc ctcctgtgtt gcccaggcct ggaattcaaa ctcctgggcc 1620
taagcaatcc tcccaccctg ggcctcccag agtgccggga ataccagggg tgaagccacc 1680
atgctctggc aaattatttt aaaataccag ggttaaaagt aaat 1724

<210> 16
<211> 4157
<212> DNA
<213> Homo sapiens

<220>
<221> misc_feature
<223> Incyte ID No: 7197873CB1

<400> 16
tgcatgtctt tattgtaggc atgagcctgt cctctgtgac gctggccagc gccctacagg 60
tcaggggtga agctctgtct gaggaggaaa tctggtcccc cctgttcctg gccgctgagc 120
agctcctgga agacctccgc aacgattcct cggactatgt ggtttgcccc tggtcagccc 180
tgctttctgc agctggaagc ctttctttcc aaggccgtgt ttctcatata gaggctgctc 240
ctttcaaggc ccctgaactg ctacagggac agagtgagga tgagcagcct gatgcatctc 300

```
agcccctgca gctctgcgag cccctgcact ccatcctgct gaccatgtgt gaagaccagc 360
ctcacaggcg gtgcacgttg cagtcggttc tggaagcttg tcgggttcat gagaaagaag 420
tgtctgtcta cccagcccct gctggtctcc acatcagaag gctggttggc ttggttctgg 480
gtaccatttc tgaggtggag aaaagagttg tggaggaaag ctcctctgtg cagcagaaca 540
gaagctacct gctcaggaag aggctgcgtg ggacaagcag cgagagccca gcggcacagg 600
ccccggagtg tctgcatcct tgcagagttt cagaaagaag cacggagacc cagagctcac 660
cagagcccca ttggagcacc ttgacacaca gtcactgcag cctccttgtt aaccgcgctc 720
ttccaggagc agatccccag gaccagcagg cgggccggag gctcagctct ggatctgtgc 780
actcggcagc agacagctca tggccaacaa ctccttctca gaggggtttt ctgcaaagaa 840
ggagcaagtt ttccaggcca gagttcatcc tgttggctgg agaggccccg atgacactac 900
atctgccggg atcggttgtg accaaaaaag ggaaatccta tttggctctc agggacctct 960
gtgtggtcct gctgaacggg cagcacctgg aggtaaaatg tgatgttgaa tcaacagtgg 1020
gagctgtctt caatgccgtg acatcctttg ccaacctcga ggaactcacc tactttggct 1080
tggcgtatat gaaaagcaaa gagttctttt tcctggacag tgaaaccaga ttgtgcaaaa 1140
tagctcctga aggctggaga gagcagcctc agaagacctc catgaatacc ttcacactct 1200
tcctgaggat aaagttcttt gtcagccact atgggctgct ccagcacagc ctgacaaggc 1260
accagtttta cctgcagctt cggaaagata tcctggagga gaggctgtac tgcaatgaag 1320
agatactgct gcagctgggg gtccttgcct tgcaggctga gtttggcaat taccctaagg 1380
agcaggtgga gagtaagcca tactttcacg ttgaagatta catcccagcg agtctgatcg 1440
agaggatgac cgctctacgg gtccaggttg aagtctcaga gatgcaccgg ctcagctctg 1500
cactgtgggg agaggatgct gagctgaagt tcttgagggt cactcagcag ctcccagaat 1560
acggtgtgct ggttcaccaa gtattctcag agaagaggag gccagaagag gagatggccc 1620
tggggatctg tgccaagggt gtcatagtct atgaagtgaa aaacaacagc agaattgcaa 1680
tgttacggtt tcagtggaga gaaaccggga agatttctac ttatcaaaaa aagttcacca 1740
tcacaagcag tgtcactggg aagaagcaca catttgtcac agattcagcc aagaccagta 1800
aatacttact ggacctctgc tcagcccagc atgggtttaa tgcacagatg ggctctgggc 1860
agccttccca tgttttattt gaccatgata agtttgtgca aatggccaat ttgagtcctg 1920
cacaccaggc ccggtctaag cctctcattt ggattcagag attgtcatgc tcagaaaacg 1980
agttgtttgt atccaggctt cagggtgctg caggaggcct gctgagtaca tcaatggata 2040
acttcaacgt ggacggcagc aaggaggctg gagcagaagg catcgggcgc agcccctgca 2100
ctggccggga gcagctgaag agtgcctgtg tgatccagaa gccaatgacc tgggactctc 2160
tctctggacc acctgttcag agcatgcatg caggctcaaa gaataatagg aggaagagct 2220
ttatagctga accgggccga gaaattgtac gtgtgacact gaaacgtgac ccacatcgtg 2280
gttttgggtt tgtcattaat gagggagagt attcaggcca agctgaccct ggcattttta 2340
tatcttctat tatacctgga ggaccagcag aaaaagcaaa aacgatcaaa ccaggagggc 2400
agatactagc cctgaatcac atcagtctgg agggcttcac attcaacatg gctgttagga 2460
tgatccagaa ttcccctgac aacatagaat taattatttc tcagtcaaaa ggtgttggtg 2520
gaaataaccc agatgaagaa aagaatagca cagccaattc tggggtctcc tctacagaca 2580
tcctgagctt cgggtaccag ggaagtttgt tgtcacacac acaagaccag gacagaaata 2640
ctgaagaact agacatggct ggggtgcaga gcttagtgcc caggctgaga catcagcttt 2700
cctttctgcc gttaaagggt gctggttctt cttgtcctcc atcacctcca gaaatcagtg 2760
ctggtgaaat ctactttgtg gaactggtta aagaagatgg gacacttgga ttcagtgtaa 2820
ctggtggcat taacaccagt gtgccatatg gtggtatcta tgtgaaatcc attgttcctg 2880
gaggaccagc tgccaaggaa gggcagatcc tacaggtgga ccgactcctg caggtggatg 2940
gagtgattct gtgcggcctc acccacaagc aggctgtgca gtgcctgaag ggtcctgggc 3000
aggttgcaag actggtctta gagagaagag tccccaggag tacacagcag tgtccttctg 3060
ctaatgacag catgggagat gaacgcacgg ctgtttcctt ggtaacagcc ttgcctggca 3120
ggccttcgag ctgtgtctca gtgacagatg gtcctaagtt tgaagtcaaa ctaaaaaaga 3180
atgccaatgg tttgggattc agtttcgtgc agatggagaa agagagctgc agccatctca 3240
aaagtgatct tgtgaggatt aagaggctct ttccggggca gccagctgag gagaatgggg 3300
ccattgcagc tggtgacatt atcctggccg tgaatggaag gtccacggaa ggcctcatct 3360
tccaggaggt gctgcattta ctgagagggg ccccacagga agtcacgctc ctcctttgcc 3420
gacccccctcc aggtgcgctg cctgagctgg agcaggaatg gcagacacct gaactctcag 3480
ctgacaaaga attcaccagg gcaacatgta ctgactcatg taccagcccc atcctggatc 3540
aagaggacag ctggagggac agtgcctccc cagatgcagg ggaaggcctg ggtctcaggc 3600
cagagtcttc ccaaaaggcc atcagagagg cacaatgggg ccaaaacaga gagagacctt 3660
gggccagttc cttgacacat tctcctgagt cccaccctca tttatgcaaa cttcaccaag 3720
aaagggatga atcaacattg gcgacctctt tggaaaagga tgtgaggcaa aactgctatt 3780
cagtttgtga tatcatgaga cttggaagat attccttctc atctcctcta accagacttt 3840
cgacagatat tttctgagca ccttctctgc atgtctgcag tgctgtgtaa aatgccctac 3900
ctttgcatgg actattcttt ctaatcaaga ggcgtgtgtg gcgaacttgg ggcagcccct 3960
ggaagtcttg ttctttgacc attacgtctg cggctgcatc accagataat gagcttcacc 4020
actcgtctgc ctcctgtgtc cttccgcggg gagtaaatgt cacttcagct tgccgcatct 4080
```

ctaaataggc aaattttcag tgctcagaaa aggacctgat cttgcacaaa gtgctttgat 4140
ggttgcctgc ttggggc 4157

<210> 17
<211> 1044
<212> DNA
<213> Homo sapiens

<220>
<221> misc_feature
<223> Incyte ID No: 6282188CB1

<400> 17
atgctgaaac accccgtgtt acctgccctg tgcctggcgc tcgtcagtct attcgccaat 60
gtttctgtgc aggccgacgc aatcgtcact tccgtccggt cccccgaatg ggcccaaccg 120
atcgacgctc actacaactt gcaccagatg acgcccacgc tctaccgcag cggcttgccg 180
gacagccgcg cgctgcctct gctggaaaaa ctgaacgttg gcaccgtcat caacttcctg 240
cccgaatccg atgacagctg gctcgccgac tccgatatca aacaagtgca gctgacgtat 300
cgcaccaacc acgtagacga ttcagatgta ttggccgcat tgcgcgcaat ccgacaggca 360
gaagccaatg gctcggtgtt gatgcactgc aagcacggct cggaccgcac cggcctgatg 420
gcggcgatgt atcgggtggt gattcaaggg tggagcaaag aggatgcgct gaacgaaatg 480
acgttgggcg ggtttggcag cagtaatggc ttcaaggacg gtgttcgcta catgatgcgc 540
gccgatatcg acaaattacg cactgccttg gccaccgggg attgcagcac cagcgcgttt 600
gcgctgtgtt cgatgaagca atggatttcc acgacaggca gtgagcagaa ggagtagaaa 660
cggatcaggc agcagcggtc cggttgaatg gacgcgccgc ctgctctggg tgtgctcagt 720
ccttttttcaa cttcggattt ggaaagaatt gcaccgcctg cacctttggg tccggcactt 780
tcagcgctga ggtattaacc cgcgtgccca attccttggg caccgacaag ccttggtcat 840
tgagcgtatc ggtgtagccg caggccacgc attcgcgatg gggcacgctg tcctcggtcc 900
acattttcaa cttatccggc tcgctgcacg cggggcacac ggccccggcg ataaattgct 960
ttttggtgat cacaggccct tcgctcatgc tgctgcatcc tcactcaggc cgctgtgacg 1020
caacagtgcg tcaatggacg gctc 1044

<210> 18
<211> 2797
<212> DNA
<213> Homo sapiens

<220>
<221> misc_feature
<223> Incyte ID No: 2182961CB1

<400> 18
catgggaaga ccccgtcttt aaaaaaaaaa aaaaatagaa ttaccatatg actcagcaat 60
tacacttttt tggatattgt attagttcat tttcatgctg ctgataaagc catacccaag 120
actgggaaga aaaagagatt taattggact tacagttcca catgcctggg gaggcctcag 180
catcacggta ggaggcaaaa ggcacttctt acatggtggc ggcaagagaa aatgaggaag 240
aagcaaaaga ggaaacccct gataaactca tcagatcatg tgagccaggc tccatgactg 300
taacttggac cacatgggtc ccaacccgct ctgaagtgca attcgggttg cagccgtcgg 360
ggcccctgcc cctccgcgcc cagggcacct tcgtcccctt tgtggacggg ggcattctcc 420
ggcggaagct ctacatacac cgagtcacgc ttcgcaagct gctgccaggg gttcagtatg 480
tttatcgctg tggcagtgcg cagggctgga gccgtcggtt ccgcttcagg gccctcaaga 540
atggggccca ctggagtccc cgtctggctg tgtttggaga cctgggggct gacaacccga 600
aggccgtccc ccggctgcgc agggacaccc agcagggcat gtatgacgcc gttctccatg 660
tgggagactt tgcctacaac ctggatcagg acaacgcccg tgttggggat aggttcatgc 720
ggctcattga acccgtggct gccagcctgc cgtacatgac atgccctggg aatcatgaag 780
aacgctacaa cttctctaac tacaaggctc gcttcagcat gccgggggat aatgaggcc 840
tgtggtacag ctgggatctg ggtcccgccc acatcatctc cttctccacc gaggtctatt 900
tctttctcca ttatggccgc cacttggtac agaggcagtt tcgctggctg gagagcgacc 960
tccagaaagc caataagaac cgggcagccc ggccgtggat catcactatg ggcaccggc 1020
ccatgtactg ctccaacgca gatctggacg actgcacacg acatgaaagc aaggtccgca 1080
aaggcctcca aggcaagctg tacgggttgg aggatctttt ctacaaatat ggagtggatc 1140
tgcagctgtg ggctcatgag cactcgtatg aacgactgtg gccaatttac aactaccagg 1200
tatttaacgg cagccgagag atgccctaca ccaacccgcg agggcctgtc cacatcatca 1260

```
caggatctgc tggctgtgag gagcggctga cgccctttgc tgtcttcccg aggccctgga 1320
gtgccgtgcg tgtgaaggag tacgggtata cgcggctgca catcctcaac gggacccaca 1380
tccacatcca gcaggtgtcg gacgaccagg atgggaagat cgtagatgat gtctgggtgg 1440
tgagacccct gtttggccgg aggatgtacc tctagggatg gcggcactct cctccagaag 1500
cctaggtttt gccgccttgg ctgctgtgac cagaaactgc ccaggcctgg gtggggagtt 1560
gggtgggccc tgactcccct gccctccaga ggccccatgt agggtacatg cagccctatg 1620
gagctggggc agctgttccc tcctggagag gtgggagtcc tggctggctg tggagggagg 1680
gcaggtgtgc gggcacagag tgacacacgg caggtttctg ctggcagggc cccaccctcc 1740
tgcatagctc tgatcgggcg aggtgcccac ggggcttcag gaatgaagag gcttaagctc 1800
tggctccatg gattctgcac atctgcgggg gatgccgctg ggcttcctcc tctcctgccc 1860
acctggcaag ggcatcgcca ggtgggcaca accgtcatga cactactcac cagcaggtgg 1920
cgtcaggggc tttttcttct gagcccggca ctgagagttg gtctgaagcc tggctccttc 1980
ttcactgctc caggactgct atgaagagtc ccttcatgcc tcagtttccc agcctggcac 2040
catcttattc gggaagagga gacgtgttaa cactcttgcc tcctagctag gacagatgac 2100
caaccgcaag agccacagac ttgccagttc cttccctctt tccttccttt ctttcccttc 2160
ttttatttat tgaatcataa tttattgagc atctaccatg tgccaggctc tgttctcagc 2220
gctggagaga cagctgtgaa tgagacagag atctcggccc tcacagagct gacatcctaa 2280
ccagagagtt ggacaaaaat cacgataaat gagttggtta aatagcgatt tgtgagtaga 2340
aaacgcaggg acggtgagag agcagtttca attttcaggg ggatctcact gagagggcaa 2400
catttgatct gaaggaggtg gggaaggagc caagtgggca gacatctggg ggaagagcat 2460
tccaggcaga ggaacagcca gtgcaaaggc cctgagacag aaatgtgcct ggccggctgg 2520
gtacagtgac tcacatctgt ggtcccagca ctttgggagg ccgaggcggg cagatcgctt 2580
gagcccagga gtttgagacc agcctgggca acacagcgaa accctttctc taccaaaaat 2640
atgaaattta actgggcatg gtggtgtgtg cctgtggtcc tagctgctcg ggaggctgag 2700
gcgggaggat ggcttgagcc caggaggttg aggctgcagt gagccatgat tgcgccactg 2760
caccccagtc tgggcaacag agcggagacc tgtctca                          2797

<210> 19
<211> 3488
<212> DNA
<213> Homo sapiens

<220>
<221> misc_feature
<223> Incyte ID No: 5119906CB1

<400> 19
tttggttcgt gtttttctct tggcatgctt tattctctgt cctgcttgtt ttttgtcctc 60
tgtcttcgtg gtttttttctc ttcgtcgtcc ttggtccttt ttgcttcact gttctttttc 120
ttgtctcgct tttttctgtc tcggtctctt ttctctctct atagtgcgcg gtgtttggcc 180
ttggttcttg tgtgcgtttt tctggctgtt cgcgcgctgc ttttgacccg tgtgtttgtc 240
cgtcagtgtt ctcgttttct gctctcctct tgtctgctcc tgctgtcttt cttttctgat 300
ggtgtctctc gtatgtctat gggttcctgt tgtccgtccc gtgtgtcttt ggtctttctt 360
ttccgtgcct cgttctccct ctcctttctg tcttgtgttt cgttctgtct ggcttttttc 420
gtctgtcctt gctttctttg tctattatct gctctttgct tgctggtcat tgttccgctg 480
tctccttgct ctctgagctt gcccgtcccc ttggttcgct cggtttaatg cgattttttt 540
tgcgggaagc ggggactgtg tctgcgggta cctcgcaatg ccctcgctca tcttgggaac 600
tgtgtctctt gtcttgcccc ttgcccagtg tctcgtgcga aatgcggggt ctccgcttac 660
agtccctcag tacattgtgg acactgataa tgtgcgttgt tcccacgcgt gcgcacgtgg 720
tcttggctcc ctcctaccca gatgtgactt tcactgcagg cgctgatttc tcaccacaga 780
tacctttctc tctgtgtttt attctcagtg ggttcagcgt gagcacagca ggaaggatgc 840
acatatttaa gcctgtgtct gtccaggcca tgtggtctgc cctgcaggtg cttcacaagg 900
cctgcgaagt ggcccggagg cacaactact tccccggggg tgtagctctc atctgggcta 960
cctactatga gagctgcatc agctccgagc agagctgcat caacgagtgg aacgccatgc 1020
aggacctgga gtctacgcgg cccgactccc ccgcgctatt tgtggacaag cccactgaag 1080
gggaaaggac cgagcgcctc atcaaagcca agctccgaag catcatgatg agccaggatc 1140
tagaaaatgt gacttccaaa gagattcgta atgaattaga gaaacagatg aattgtaact 1200
tgaaggaact caaggaattt atagacaatg agatgctact tatcttggga cagatggaca 1260
agccctccct tatcttcgat catctttatc tcggctctga atggaatgca tccaatctgg 1320
aggaactgca gggctcaggg gttgattaca ttttaaatgt taccagagaa atcgataatt 1380
tttttcctgg cttatttgca tatcataaca tccgagtcta cgatgaagag accacagacc 1440
tcctcgccca ctggaatgaa gcgtatcatt ttataaacaa agcgaagagg aaccattcca 1500
agtgcctggt gcattgcaaa atgggcgtga gtcgctcggc ctccacagtc atagcctatg 1560
```

```
caatgaagga attcggctgg cctctggaaa aagcatataa ctatgtaaag cagaagcgca 1620
gcatcacgcg ccccaacgcg ggctttatga ggcagctgtc tgagtatgaa ggcatcttgg 1680
atgcaagcaa acagcggcac aacaagctgt ggcgtcagca gacagacagc agcctccagc 1740
agcctgtgga tgaccctgca ggacctggcg acttcttgcc agagacccca gatggcaccc 1800
cggaaagcca gctgcccttc ttggatgatg ccgcccagcc cggcttaggg cccccctcc 1860
cctgctgttt ccggcgactc tcagaccccc ttctgccttc ccctgaggat gaaactggca 1920
gcttggtcca cctggaggat ccggagaggg aggctctgtt ggaggaagct gctccacctg 1980
cagaggtgca caggccggcc agacagcccc agcaaggttc cggactctgt gagaaggatg 2040
tgaagaagaa actagagttt gggagtccca aaggtcggag cggctccttg ctgcaggtgg 2100
aggagacgga aagggaggag ggcctgggag cagggaggtg gggcagcttt ccaacccagc 2160
tcgatcaaaa cctgctcaac tcggagaacc taaacaacaa cagcaagagg agctgtccca 2220
acggcatgga ggatgatgct atatttggga tccttaacaa agtgaagcct tcctataaat 2280
cctgtgccga ctgcatgtac cctacagcca gcggggctcc tgaggcctcc agggagcgat 2340
gtgaggaccc caatgctccc gccatctgca cccagccagc cttcctaccc cacatcacgt 2400
cctcccctgt ggcccacttg gccagcaggt cccgtgttcc ggagaagcca gcctctggcc 2460
caaccgaacc tcccccgttc ctaccaccag caggctccag gagggcagac accagtggcc 2520
ctggggctgg agctgcccta gagccaccag ccagcctttt ggaaccttcc agagagaccc 2580
caaaagtcct gccaaagtcc ctcctttga agaattctca ctgtgataag aaccctccca 2640
gcacagaagt ggtaataaag gaagaatcgt cacccaagaa agatatgaag ccagccaagg 2700
acctgaggct tctgttcagt aatgaatctg agaagccgac aaccaacagc tacctgatgc 2760
agcaccagga gtccatcatt cagctgcaga aggcaggctt ggtccgcaag cacaccaaag 2820
agctagagcg gctgaagagc gtgcctgcag acccagcacc tccctccagg gatggccctg 2880
ccagcaggct ggaggccagc atccccgagg agagccagga tccagccgcg ctccacgagc 2940
tgggccccct ggttatgccc agccaggccg ggagtgatga gaagtcagag gccgcccccg 3000
cttcattgga aggaggctca ctgaagagcc cccctccttt cttctaccgc ctggaccaca 3060
ccagtagttt ctcaaaagac tttctgaaga ccatctgcta cacccccacc tcctcttcca 3120
tgagctccaa cctgacccgg agctccagca gcgatagcat ccacagtgtc cgtgggaagc 3180
ccgggctggt gaagcagcgg acacaggaga ttgagacccg gctccggctg gcgggcctca 3240
ccgtctcttc cccactgaag cgctcacact ctcttgccaa gctggggagt ctcaccttct 3300
caacggaaga cctgtccagt gaggctgacc cgtccaccgt cgctgactcc caggacacca 3360
ctttgagtga atcttccttc ttgcatgagc cccagggaac cccgagggac ccagctgcaa 3420
cctccaaacc atcagggaaa cccgccccag aaaacttaaa gagcccttcg tggatgagca 3480
aaagctga                                                          3488

<210> 20
<211> 1522
<212> DNA
<213> Homo sapiens

<220>
<221> misc_feature
<223> Incyte ID No: 4022502CB1

<400> 20
tgcggtgcca gcggagggca cggcccggcg tggcagcggc ggcggacgcg gccccgggca 60
caccatggcc gagctgctgc ggagcctgca ggattcccag ctcgtcgccc gcttccagcg 120
ccgctgcggg ctcttccccg ctccggatga aggcccccgg gagaacggcg cggaccccac 180
ggagcgcgcg gcgcgggtcc ccggggtcga gcatctcccc gcagccaacg gcaagggcgg 240
cgaggctccg gccaacgggc tgcgcagagc cgcggcgccg gaggcttatg tacagaagta 300
cgtcgtgaag aattatttct actattacct attccaattt tcagctgctt tgggccaaga 360
agtgttctac atcacgtttc ttccattcac tcactggaat attgaccctt atttatccag 420
aagattgatc atcatatggg ttttggtgat gtatattggc caagtggcca aggatgtctt 480
gaagtggccc cgtccctcct cccctccagt tgtaaaactg gaaaagagac tgatcgctga 540
atatggaatg ccatccaccc acgccatggc ggccactgcc attgccttca ccctccttat 600
ctctactatg gacagatacc agtatccatt tgtgttggga ctggtgatgg ccgtggtgtt 660
ttccaccttg gtgtgtctca gcaggctcta cactgggatg catacggtcc tggatgtgct 720
gggtggcgtc ctgatcaccg cactcctcat cgtcctcacc taccctgcct ggaccttcat 780
cgactgcctg gactcggcca gcccccctctt ccccgtgtgt gtcatagttg tgccattctt 840
cctgtgttac aattaccctg tttctgatta ctacagccca acccgggcgg acaccaccac 900
cattctggct gccggggctg gagtgaccat aggattctgg atcaaccatt tcttccagct 960
tgtatccaag cccgctgaat ctctccctgt tattcagaac atcccaccgc tcaccaccta 1020
catgttagtt ttgggtctga ccaaatttgc agtgggaatt gtgttgatcc tcttggttcg 1080
tcagcttgta caaaatctct cactgcaagt attatactca tggttcaagg tggtcaccag 1140
```

```
gaacaaggag gccaggcgga gactggagat tgaagtgcct tacaagtttg ttacctacac 1200
atctgttggc atctgcgcta caacctttgt gccgatgctt cacaggtttc tgggattacc 1260
ctgagtctca aacagttgga aactagccca ctggacatga aagccaagac ataggaaagt 1320
tattggtagg caaatcttga caacttattt ttctttaaca acaacaaaaa gtcatacggc 1380
tgtcttgcta ctaccagata aatgatgctg ctgtgtgaaa ggaagaactg tctcatagcg 1440
gtcattggtc gtccgtggtg gttggttgtg ctacagttga acccaggcta aagaccataa 1500
tccggatctt taaaggcaca ca                                          1522

<210> 21
<211> 1393
<212> DNA
<213> Homo sapiens

<220>
<221> misc_feature
<223> Incyte ID No: 4084356CB1

<400> 21
atgagagcgt ggatccctgg gtgggttggg cggccgcacg gggtgccga ggcgtctggg 60
ggcctgcgct tcggggcgag cgcagcgcaa ggctggcgcg cgcgcatgga ggatgctcac 120
tgcacttggc tttcgttacc tggtctgccc ccgggctggg ccttgtttgc cgtcctcgac 180
ggccacggtg gggctcgagc tgcccgcttc ggtgcacgcc atttgccagg ccatgtgctc 240
caggagctgg gcccggagcc tagcgagccc gagggcgtgc gcgaggcgct gcgccgagcc 300
ttcttgagcg ccgacgagcg ccttcgctcc ctctggcccc gcgtggaaac gggcggcttc 360
acggccgtag tgttgctggt ctccccgcgg tttctgtacc tggcgcactg cggtgactcc 420
cgcgcggtgc tgagccgcgc tggcgccgtg gccttcagca cagaggacca ccggcccctt 480
cgaccccggg aacgcgagcg catccacgcc gctggcggca ccatccgccg ccgccgcgtc 540
gagggctctc tggccgtgtc gcgagcgttg ggcgacttta cctacaagga ggctccgggg 600
aggccccccg agctacagct cgtttctgcg gagccagagg tggccgcact ggcacgccag 660
gctgaggacg agttcatgct cctggcctct gatggcgtct gggacactgt gtctggtgct 720
gccctggcgg gactggtggc ttcacgcctc cgcttgggcc tggccccaga gcttctctgc 780
gcgcagctgt tggacacgtg tctgtgcaag ggcagcctgg acaacatgac ctgcatcctg 840
gtctgcttcc ctggggcccc taggccttct gaggaggcga tcaggaggga gctagcactg 900
gacgcagccc tgggctgcag aatcgctgaa ctgtgtgcct ctgctcagaa gccccccagc 960
ctgaacacag ttttcaggac tctggcctca gaggacatcc cagatttacc tcctggggga 1020
gggctggact gcaaggccac tgtcattgct gaagtttatt ctcagatctg ccaggtctca 1080
gaagagtgcg gagagaaggg gcaggatggg gctgggaagt ccaaccccac gcatttgggc 1140
tcagccttgg acatggaggc ctgacagctg ttgtcctttg gggatccttt gcttctctgg 1200
ggcctcaaca gaactaaaga agaaaaccga ccctttcccc aactacatgt accagcggaa 1260
ggaaggaagg ccaatgtagg aacccaaaat gcttatttct tcttctctta cttccctctc 1320
acagaaaagt cttacgaatg gggaaattcc accaacatcc agaccaaaaa gaaaaaagcc 1380
caaatcgaaa aaa                                                    1393

<210> 22
<211> 1430
<212> DNA
<213> Homo sapiens

<220>
<221> misc_feature
<223> Incyte ID No: 1740204CB1

<400> 22
ccagagcacc gggcacggcc ttcaatgggc gaggacacgg acacgcggaa aattaaccac 60
agcttcctgc gggaccacag ctatgtgact gaagctgaca tcttctctac cgttgagttc 120
aaccacacgg gagagctgct ggccacaggt gacaagggcg gccgggtcgt catcttccag 180
cgggaaccag agagtaaaaa tgcgccccac agccagggcg aatacgacgt gtacagcact 240
ttccagagcc acgagccgga gtttgactat ctcaagagcc tggagataga ggagaagatc 300
aacaagatca agtggctccc acagcagaac gccgcccact cactcctgtc caccaacgat 360
aaaactatca aattatggaa gattaccgaa cgagataaaa ggcccgaagg atacaacctg 420
aaggatgaag aggggaaact taaggacctg tccacggtga cgtcactgca ggtgccagtg 480
ctgaagccca tggatctgat ggtggaggtg agccctcgga ggatctttgc caatggccac 540
acctaccaca tcaactccat ctccgtcaac agtgactgcg agacctacat gtcggcggat 600
```

```
gacctgcgca tcaacctctg gcacctggcc atcaccgaca ggagcttcaa catcgtggac 660
atcaagccgg ccaacatgga ggaccttacg gaggtgatca cagcatctga gttccatccg 720
caccactgca acctcttcgt ctacagcagc agcaagggct ccctgcggct ctgcgacatg 780
cgggcagctg ccctgtgtga caagcattcc aagctctttg aagagcctga ggaccccagt 840
aaccgctcat tcttctcgga aatcatctcc tccgtgtccg acgtgaagtt cagccacagc 900
ggccgctaca tgctcacccg ggactacctt acagtcaagg tctgggacct gaacatggag 960
gcaagaccca tagagaccta ccaggtccat gactaccttc ggagcaagct ctgttccctg 1020
tacgagaacg actgcatttt cgacaagttt gaatgtgcct ggaacgggag cgacagcgtc 1080
atcatgaccg ggcctacaa caacttcttc cgcatgttcg atcggaacac caagcgggac 1140
gtgaccctgg aggcctcgag ggaaagcagc aagccccggg ctgtgctcaa gccacggcgc 1200
gtgtgcgtgg ggggcaagcg ccggcgtgat gacatcagtg tggacagctt ggacttcacc 1260
aagaagatcc tgcacacggc ctggcacccg gctgagaaca tcattgccat cgccgccacc 1320
aacaacctgt acatcttcca ggacaaggta aactctgaca tgcactaggt atgtgcagtt 1380
cccggcccct gccacccagc ctcatgcaag tcatccccga catgaccttc           1430

<210> 23
<211> 3102
<212> DNA
<213> Homo sapiens

<220>
<221> misc_feature
<223> Incyte ID No: 7483804CB1

<400> 23
ggcggtcggg cgagggagcg cgcacggagc gcgggacgga gcgccaggcg gacggaccga 60
aggacggagg caccgaagga cggacgcccc cgcacacgca gacgcacaga gctcggcgcg 120
gcccccggtc gcatacacac tggcacagac acaagcaggg acacacgcag acacacgcac 180
actcgcgcgc gcatcctccc gccagcctgc ccgcctgctc gccggcgccc ggagcccgct 240
ctggccgctg gatgatctga agctgccctc tctccttcat ttatatcacc agcttgcttt 300
ttgctgagaa agcttcctgc cctggaagat ggcacccttc cccatccaga caccttggga 360
atgaattatg agggagccag gagtgagaga gagaaccacg ctgctgatga ctccgaggga 420
ggggccctgg acatgtgctg cagtgagagg ctaccgggtc tcccccagcc gatagtgatg 480
gaggcactgg acgaggctga agggctccag gactcacaga gagagatgcc gccacccccct 540
cctccctcgc cgccctcaga tccagctcag aagccaccac ctcgaggcgc tgggagccac 600
tccctcactg tcaggagcag cctgtgcctg ttcgctgcct cacagttcct gcttgcctgt 660
ggggtgctct ggttcagcgg ttatggccac atctggtcac agaacgccac aaacctcgtc 720
tcctctttgc tgacgctcct gaaacagctg gaacccacgg cctggcttga ctctgggacg 780
tggggagtcc ccagtctgct gctggtcttt ctgtccgtgg gcctggtcct cgttaccacc 840
ctggtgtggc acctcctgag gacaccccca gagccaccca ccccactgcc ccctgaggac 900
aggcgccagt cagtgagccg ccagccctcc ttcacctact cagagtggat ggaggagaag 960
atcgaggatg acttcctgga cctcgacccg gtgcccgaga ctcctgtgtt tgattgtgtg 1020
atggacatca agcctgaggc tgaccccacc tcactcaccg tcaagtccat gggtctgcag 1080
gagaggaggg gttccaatgt ctccctgacc ctggacatgt gcactccggg ctgcaacgag 1140
gagggctttg gctatctcat gtccccacgt gaggagtccg cccgcgagta cctgctcagc 1200
gcctcccgtg tcctccaagc agaagagctt catgaaaagg ccctggaccc tttcctgctg 1260
caggcggaat tctttgaaat ccccatgaac tttgtggatc cgaaagagta cgacatccct 1320
gggctggtgc ggaagaaccg gtacaaaacc atacttccca accctcacag cagagtgtgt 1380
ctgacctcac cagaccctga cgaccctctg agttcctaca tcaatgccaa ctacatccgg 1440
cctggacttg gctggccgca gggctatggt ggggaggaga aggtgtacat cgccactcag 1500
ggacccatcg tcagcacggt cgccgacttc tggcgcatgg tgtggcagga gcacacgccc 1560
atcattgtca tgatcaccaa catcgaggag atgaacgaga aatgcaccga gtattggccg 1620
gaggagcagg tggcgtacga cggtgttgag atcactgtgc agaaagtcat tcacacggag 1680
gattaccggc tgcgactcat ctccctcaag agtgggactg aggagcgagg cctgaagcat 1740
tactggttca catcctggcc cgaccagaag accccagacc gggccccccc actcctgcac 1800
ctggtgcggg aggtggagga ggcagcccag caggaggggc cccactgtgc ccccatcatc 1860
gtccactgca gtgcagggat tgggaggacc ggctgcttca ttgccaccag catctgctgc 1920
cagcagctgc ggcaggaggg tgtagtggac atcctgaaga ccacgtgcca gctccgtcag 1980
gacaggggcg gcatgatcca gacatgcgag cagtaccagt ttgtgcacca cgtcatgagc 2040
ctctacgaaa agcagctgtc ccaccagtcc ccagaatgac tgcgcttctc ctacaaggtt 2100
ctctgggcac tgcccagcct gagtctcggc cctcacccag ggccctgcct cgggtcctgg 2160
gcctgctccc cgcttcctcc ccttcagtca gctccctctg tcctctgtca gcctggcctg 2220
acccctaccc tccagcattg ctcttcctac tgtacatatt ggggagtggg gggcagggtc 2280
```

```
gggaagggac atgccaggcc aggcctgggg ccccggggcc tgacccacac cacgcagacc 2340
ccgggctcca gtttttaacg atggttccat caatacctga tccagaatgt ttccgtgcta 2400
cactttgtgt cctgctgcaa tgtgttctgt ctgtccatcc atctctgccc tctgtaccgg 2460
acactgtgtc tcctcagcca ggaaggggta atgagctcca gcccctaagc aaccggactt 2520
gcctgcctcg gcctcacccg cacttctccc aaaaggcaga tgacggggag ttaggcatgg 2580
ggagctccag aaggtcacca gagagctttc agctgaggga gagttctcta ggttggagtg 2640
ggcatcacag ccagggtggc ctctgggtgt cagatgctct caggagggtg cccagcctgt 2700
gaggcactgg caaggtaggg ggcagatggg gcatggagaa cccagaggat ctaggccctg 2760
ttggggaggg gaggggagct caaggtttgg gtggggactc agcccagatc tacgtgagac 2820
atttttctgt gtcactgtgg gaaagccttc ccagaagtct cactgcgtgt tgctctgcgt 2880
gtgttcccat gtccatgcgt gtgttgagag cccatcagga gggcatgcat gactctttgg 2940
caacatgtat tatcttggag ccacgtgttt ttattgctga ctttaaatat ttatcccacg 3000
gcagacagag acatttggtg tctttttata attcgctcgt ggtcattgaa tagagcaata 3060
aacggagcat tttgagcaaa accaaaaaaa aaaaaagggg cg                    3102

<210> 24
<211> 5612
<212> DNA
<213> Homo sapiens

<220>
<221> misc_feature
<223> Incyte ID No: 7483934CB1

<400> 24
cgaaagagtg actccaagtc agccctaaac acccataccc accccacctc ccgtgcagct 60
ctgacaggcc aagcctgggc gctagaaccc cgcgccagaa ggccgggcca cacactccag 120
cgtacacgcg cggacacgtg ccggtgtcca gagccggtgg gagaggcgct gtcaccatct 180
gacaggagag gaaggtggag cgcagagaag tcagggtcac gtgtgggccc gaccctcatg 240
gccctgagca aagggctgcg gctgctgggg cgcctggggg ccgagggggа ctgtagcgtg 300
ctgctggagg cgcgcggccg cgacgactgc ctgctgttcg aggccggcac ggtggccacg 360
ctggacgact gcctgctgtt cgaggccggc acggtggcca cgctggctcc agaagaaaag 420
gaagtcatta aaggacagta tggcaagctc acggacgcgt acggctgcct gggggagctg 480
aggctgaaat ctggtggcac gtctctgagc ttcctggtgt tggtgacagg ctgcacatct 540
gtgggcagaa ttccagatgc tgaaatctac aaaatcactg ccactgactt ttaccctctt 600
caggaagagg ccaaggagga ggaacgcctc atagctttga agaaaatcct cagctcgggg 660
gtgttctatt tctcatggcc aaacgatggg tctcgctttg acctgactgt ccgcacgcag 720
aagcaggggg atgacagctc tgaatggggg aactccttct tctggaacca gctgttgcac 780
gtgcccttga ggcagcacca ggtgagctgc tgtgactggc tgctgaagat catctgcggg 840
gtggtcacca tccgcaccgt gtatgcctcc cacaagcagg ccaaggcctg cctcgtctct 900
cgcgttagct gtgagcgcac aggcactcgc ttccacaccc gtggcgtgaa cgacgacggc 960
catgtgtcca acttcgtgga gacagagcag atgatttaca tggacgatgg agtgtcatct 1020
tttgtccaga tcagaggctc cgttccgctg ttctgggaac agccagggct tcaggtaggt 1080
tcccatcatc tgagactcca caaaggcctg gaagccaatg cccctgcttt cgacaggcac 1140
atggtgcttc tgaaggagca gtacgggcag caggtggtcg tgaaccttct gggaagcaga 1200
ggcggagagg aggtgctcaa cagagccttc aagaagctgc tctgggcttc ttgccacgcg 1260
ggcgacacgc ctatgatcaa ttttgacttc catcagtttg ccaaaggtgg gaagctagag 1320
aaattggaga ccctcttgag gccacagtta aagctgcact gggaagactt cgatgtgttc 1380
acaaaggggg agaacgtcag tccacgtttt cagaaaggca ctttgcggat gaactgtctt 1440
gactgcctgg accgaaccaa cactgtgcag agcttcatcg cgctcgaggt cctgcatctg 1500
cagctcaaga ccctggggct gagttcaaaa cccatcgttg accgctttgt ggagtccttc 1560
aaagccatgt ggtctctgaa tggccacagc ctgagcaagg tgttcacagg cagcagagcc 1620
ctggaaggga aggccaaggt ggggaagctg aaggatggag cccggtccat gtctcgaacc 1680
atccagtcca acttcttcga cggggtgaag caggaggcca tcaagctgct gctggttggg 1740
gacgtctacg gcgaggaggt ggcagacaaa gggggcatgc tgctggacag cacggcgctc 1800
ctggtgactc ccaggatcct gaaagctatg actgagcgtc agtccgaatt cacaaatttc 1860
aagcggatcc ggattgctat ggggacctgg aacgtgaacg gaggaaagca gttccggagc 1920
aacgtgctca ggacggcgga gctgacagac tggctgctcg actcgcccca gctctcggga 1980
gctaccgact cccaggatga cagcagccca gctgacatat ttgctgtggg gtttgaagag 2040
atggtggaat tgagcgcagg gaatattgtc aatgccagta ctaccaacaa gaagatgtgg 2100
ggtgaacagc ttcagaaagc catctcacgc tctcatagat acattctgtt gacttcggca 2160
cagctggtgg gcgtctgtct ttatatcttt gtacgtccat accatgtccc gttcatcagg 2220
gacgtagcca tcgacacagt gaagacgggc atgggggggca aggcggggaa caagggcgcc 2280
```

```
gtcggcatcc gcttccagtt ccacagcacc agcttctgct tcatatgtag tcacctgacg 2340
gccgggcagt cccaggtgaa ggagcggaat gaagactaca aggagatcac ccagaaactc 2400
tgcttcccaa tggggagaaa tgtttttttct catgattatg tattttggtg tggcgatttc 2460
aactaccgca ttgatcttac ttatgaagaa gtcttctatt ttgttaaacg ccaagactgg 2520
aagaaacttc tggaatttga tcaactacag ctacagaaat caagtggaaa aatttttaag 2580
gactttcacg aaggagccat taactttgga cccacctaca agtatgacgt tggctcagcc 2640
gcctacgata caagcgacaa atgccgcacc cccgcctgga cagacagggt gctgtggtgg 2700
aggaagaaac atccctttga taaaacagct ggagaactca accttctaga cagtgatcta 2760
gatgttgaca ccaaagtcag acacacctgg tctcctggtg ccctgcagta ttatggtcgt 2820
gcggagctac aagcgtctga tcacagacct gtgctggcga tcgtggaggt ggaagttcag 2880
gaagtcgatg tgggtgctcg ggagagggtt ttccaggaag tgtcctcctt ccagggcccc 2940
ctggatgcca ctgttgtagt aaaccttcaa tcaccgacct tagaagagaa aaacgagttt 3000
ccagaggacc tgcgtactga gctcatgcag accttgggga gttatgggac aattgttctt 3060
gtcaggatca accaagggca gatgctggta acttttgcag acagtcactc ggctctcagt 3120
gtcctggacg tggacggtat gaaggtgaaa ggcagagcag tgaagattag accgaagacc 3180
aaggactggc tgaaaggttt gcgagaggag atcattcgga aacgagacag catggccccc 3240
gtgtctccca ctgccaactc ctgtttgctg gaggaaaact ttgacttcac aagtttggac 3300
tatgagtcag aaggggatat tcttgaagac gatgaagact acttggtgga tgaattcaat 3360
cagcctgggg tctcggacag tgaactcggg ggagacgacc tctctgatgt ccccggcccc 3420
acagcactgg ctcctcccag caagtcacct gctctcacca aaaagaagca gcatccaacg 3480
tacaaagatg acgcggacct ggtggagctc aagcgggagc tggaagccgt cggggagttc 3540
cgccaccgtt ctccgagcag gtctctgtcg gtccccaacc gacctcggcc acctcaaccc 3600
ccgcagagac ccccccctcc aaccggttta atggtgaaaa agtcggcttc agatgcgtcc 3660
atctcctccg gcacccatgg gcagtattca attttgcaga cggcaagact tctaccagga 3720
gcacctcagc aacctcccaa ggctcggact ggaataagta aaccttataa tgtcaagcag 3780
atcaaaacca ccaatgccca ggaggcagaa gcagcaatcc ggtgtctcct ggaagccaga 3840
ggaggtgcct ccgaagaagc cctaagtgcc gtggccccaa gggaccttga agcatcctct 3900
gaaccagagc ccacaccggg ggcagccaaa ccagagaccc cacaggcgcc cccactcctt 3960
ccccgtcggc ccccacccag agttcctgcc atcaagaagc caaccttgag aaggacagga 4020
aagcccctgt caccggaaga acagtttgag caacagactg tccattttac aatcgggccc 4080
ccggagacaa gcgttgaggc ccctcctgtc gtgacagccc ctcgagtccc tcctgttccc 4140
aaaccaagaa catttcagcc tgggaaagct gcagagaggc caagccacag gaagccagca 4200
tcagacgaag cccctcctgg ggcaggagcc tctgtgccac cacctctgga ggcgccgcct 4260
cttgtgccca aggtaccccc gaggaggaag aagtcagccc ccgcagcctt ccacctgcag 4320
gtcctgcaga gcaacagcca gcttctccag ggcctcactt acaatagcag tgacagcccc 4380
tctgggcacc cacctgccgc gggcaccgtc ttcccacaag ggactttctt cagcacttca 4440
tctgctacaa gccccgacag cgatggcacc aaagcgatga agccagaggc agccccactt 4500
cttggtgatt atcaggaccc cttctggaac cttcttcacc accctaaact gttgaataac 4560
acttggcttt ctaagagctc agaccctttg gactcaggaa ccaggagccc caaaagagat 4620
cccatagacc cagtgtcagc tggcgcttca gctgccaagg cagagctgcc accagatcat 4680
gaacacaaaa ccttaggtca ctgggtgaca atcagtgacc aagaaaagag gacagcactg 4740
caggtgtttg acccactggc aaaaacatga ctgagcagct ttgaaggctg cagtcctata 4800
gaatgcatac cttcctccct ctagacatcc ctccaccaga agagacatct atttaaaggc 4860
acactggcca aaacgtttgt gcatctgtca ctctcgtgta gtttacaaaa atcgtgtctc 4920
ttattcagta agatggttac tcagccacca aaatatattt cactcaaggc ttgtacatct 4980
gaagtttgct cttcaaggaa tgggaacctt cctgttaaat tcggtgtatg gattttaaga 5040
aaggaatcta gccaatgagg tccaagaagt tctcacccat tgaattttta aatggctgtt 5100
cagttcatgt tgtacgtgat ggagatttgt cttttgtttt atttgcattt tacagatttg 5160
gtataacatt ttggggagcc acctgaaggt tgatgtataa agtaaggatt agagaaagag 5220
gtcgttgtga ccattagtag ctgtcctggc ccacttaaac aaggttacaa aaaatcagag 5280
tcggaagcag ccaaataggt caacctaatg actagactgt acattcccat gagccttcat 5340
gtttaagtgt gtacatgtgc gttaaccttg atgatgcgtg aatcccgagg gagccggtgg 5400
catacaccgt tagcttaacc ttagcttaaa ctagctgaag gctcctgtgc catgtcttag 5460
acattgcatg ccctatcaat tactataatc ctgagccatg gtgtgctact gaaaccaatt 5520
tttatccacc atctagtcct tattaaatga aacctcacgg atcctttgtt ccgcttatat 5580
tccatgcata ccacataaaa gcacacagtg cg                               5612
```